



(10) **DE 10 2023 124 878 A1** 2024.03.21

(12) **Offenlegungsschrift**

(21) Aktenzeichen: **10 2023 124 878.2**

(22) Anmeldetag: **14.09.2023**

(43) Offenlegungstag: **21.03.2024**

(51) Int Cl.: **G06F 16/335 (2019.01)**

G06F 40/35 (2020.01)

G06F 3/16 (2006.01)

(30) Unionspriorität:

17/947,491

19.09.2022

US

(71) Anmelder:

NVIDIA Corporation, Santa Clara, CA, US

(74) Vertreter:

Kraus & Lederer PartGmbB, 80539 München, DE

(72) Erfinder:

Xu, Peng, Santa Clara, CA, US; Patwary, Mostofa, Santa Clara, CA, US; Shetty, Rajath, Santa Clara, CA, US; Pathak, Niraj Lalit, Santa Clara, CA, US; Kumar, Ratin, Santa Clara, CA, US; Catanzaro, Bryan, Santa Clara, CA, US; Shoeybi, Mohammad, Santa Clara, CA, US

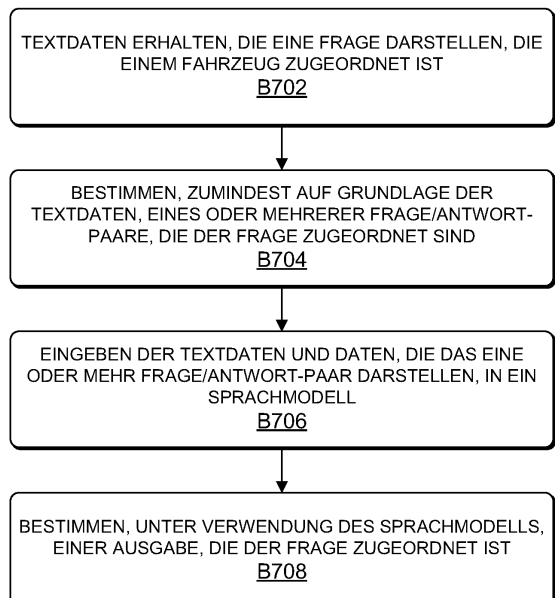
Prüfungsantrag gemäß § 44 PatG ist gestellt.

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.

(54) Bezeichnung: **DIALOGSYSTEME UNTER VERWENDUNG VON WISSENSBASEN UND SPRACHMODELLEN FÜR KRAFTFAHRZEUGSYSTEME UND -ANWENDUNGEN**

700

(57) Zusammenfassung: In verschiedenen Beispielen werden Systeme und Verfahren beschrieben, die Dialogsysteme verwenden, die verschiedenen Maschinensystemen und -anwendungen zugeordnet sind. Zum Beispiel können die Systeme und Verfahren Textdaten empfangen, die Sprache darstellen, wie eine Frage, die einem Fahrzeug oder einem anderen Maschinentyp zugeordnet ist. Die Systeme und Verfahren verwenden dann ein oder mehrere Abfragesysteme, um ein oder mehrere Frage/Antwort-Paare abzurufen, die den Textdaten und/oder den Textdaten zugeordneten Kontextinformationen zugeordnet sind. In manchen Beispielen sind die Kontextinformationen einer Wissensbasis zugeordnet, die dem Fahrzeug zugeordnet ist oder diesem entspricht. Die Systeme und Verfahren erzeugen dann eine Eingabeaufforderung unter Verwendung der Textdaten, des/der Frage/Antwort-Paares/e und/oder der Kontextinformationen. Darüber hinaus bestimmen die Systeme und Verfahren unter Verwendung eines Sprachmodells bzw. von Sprachmodellen und zumindest auf Grundlage der Eingabeaufforderung eine den Textdaten zugeordnete Ausgabe. Die Ausgabe kann zum Beispiel Informationen umfassen, welche die dem Fahrzeug zugeordnete Frage beantworten.



Beschreibung**STAND DER TECHNIK**

[0001] Fahrzeuge können mit digitalen Assistenten oder Gesprächsassistenten ausgestattet sein, die verschiedene Aufgaben erfüllen können, wie zum Beispiel die Bereitstellung von Informationen für Fahrgäste auf Anfrage. Damit ein Gesprächsassistent in einem Fahrzeug arbeiten kann, kann er mit einer Reihe von Antworten auf eine Reihe von Fragen, die häufig von Fahrgästen gestellt werden, vorgeladen werden. Die Fragen und Antworten können zum Beispiel anhand eines Handbuchs des Originalgeräteherstellers (OEM) für ein Fahrzeugmodell ermittelt werden. In Erwartung der Frage eines Fahrers oder Beifahrers nach dem empfohlenen Reifendruck für das Fahrzeug kann der Gesprächsassistent mit Frage/Antwort-Paaren vorgeladen werden, d.h. mit spezifischen Antworten, die Fragen zugeordnet sind. Ein Beispiel wäre die Frage „Welchen Reifendruck sollte ich für die Vorderreifen verwenden“ oder „Was ist der empfohlene Reifendruck“. In einem solchen Beispiel kann der Gesprächsassistent mit einer vordefinierten Antwort antworten, wie zum Beispiel „Der empfohlene Reifendruck beträgt sechsenddreißig Pfund pro Quadratzoll“, wobei diese Antwort dem OEM-Handbuch entnommen und/oder unter Verwendung des OEM-Handbuchs erstellt wurde.

[0002] Da diese Gesprächsassistenten jedoch mit bestimmten Frage/Antwort-Paaren vorgeladen sind, können verschiedene Probleme auftreten. Ein erstes Beispiel: Ein Gesprächsassistent ist möglicherweise nicht in der Lage, eine von einem Fahrgast gestellte Frage richtig zu interpretieren und/oder eine falsche Antwort zu geben, wenn die Frage des Fahrgastes nicht genau genug mit einer der vorinstallierten Fragen übereinstimmt. So kann der Gesprächsassistent zum Beispiel eine vom Fahrgast gestellte Frage nicht richtig interpretieren, wenn die Frage lautet: „Müssen Sie den empfohlenen Reifendruck einhalten?“, da diese Frage möglicherweise nicht für den Gesprächsassistenten vorgeladen wurde. Da diese Gesprächsassistenten mit Fragen und Antworten vorgeladen sind, die einem OEM-Handbuch zugeordnet sind, müssen die Fahrzeughersteller unter Umständen Fragen und Antworten für jeden Fahrzeugtyp (zum Beispiel jedes Fahrzeugjahr, jedes Fahrzeugmodell usw.) erstellen. So kann zum Beispiel ein OEM-Handbuch für ein bestimmtes Fahrzeugmodell und/oder -jahr andere Informationen enthalten (zum Beispiel andere empfohlene Komponentenparameter, andere Merkmale, andere Wartungspläne usw.) als ein anderes OEM-Handbuch für ein anderes Fahrzeugmodell und/oder -jahr. Darüber hinaus ist die Erzeugung dieser Frage-Antwort-Paare für umfassende Szenarien oder mehrere Bereiche äußerst arbeitsintensiv und lässt sich unter Umständen nur schwer oder gar

nicht für verschiedene Fahrzeugmarken und -modelle oder für andere Anwendungsfälle und Kontexte skalieren oder anpassen.

DARSTELLUNG

[0003] Ausführungsformen der vorliegenden Offenbarung beziehen sich auf Dialogsysteme für Automobilsysteme und -anwendungen. Es werden Systeme und Verfahren offenbart, die Audiodaten (und/oder den Audiodaten entsprechende Textdaten) erzeugen und/oder empfangen, die Sprache von einem Benutzer darstellen, wobei die Sprache eine Frage enthalten kann, die einem Fahrzeug oder einer anderen Maschine (zum Beispiel einem autonomen oder halbautonomen Fahrzeug, einer Baumaschine, einer Landschaftsbauausrüstung, einem Lagerfahrzeug, einem Flugzeug, einem wasserbasierten Fahrzeug usw.) zugeordnet ist. Die Systeme und Verfahren können dann eine oder mehrere Techniken verwenden, um Informationen abzurufen, die einem Kontext der Sprache zugeordnet sind. In einem ersten Beispiel können die Systeme und Verfahren ein oder mehrere Abfragesystem(e) verwenden, um ein oder mehrere Frage/Antwort-Paare abzurufen, die der Sprache zugeordnet sind, zum Beispiel aus einer oder mehreren Datenbanken. In einem zweiten Beispiel können die Systeme und Verfahren das (die) Abfragesystem(e) verwenden, um kontextbezogene Informationen in Bezug auf die Sprache abzurufen, zum Beispiel kontextbezogene Informationen aus einer (festen oder dynamischen) textbasierten Wissensbasis - zum Beispiel einem Handbuch, einem Fahrzeughandbuch, einem Maschinenhandbuch, einem Dokument usw. - die in der/den Datenbank(en) gespeichert ist. In jedem dieser Beispiele können die Systeme und Verfahren der vorliegenden Offenbarung Daten, welche die Sprache darstellen, Daten, welche die dem Kontext zugeordneten Informationen darstellen, und/oder andere Daten in ein oder mehrere Sprachmodelle (zum Beispiel ein großes Sprachmodell) eingeben. Das (die) Sprachmodell(e) kann (können) dann die Daten verarbeiten und auf Grundlage der Verarbeitung Daten ausgeben, die der Sprache zugeordnet sind. Enthält die Sprache zum Beispiel die dem Fahrzeug zugeordnete Frage, können das/die Sprachmodell(e) Informationen (zum Beispiel eine Antwort) ausgeben, die der Frage zugeordnet sind. Die Systeme und Verfahren können dann die Informationen dem Benutzer bereitstellen.

[0004] Im Gegensatz zu aus dem Stand der Technik bekannten Systemen, wie jenen, die oben beschrieben wurden, verwenden die aktuellen Systeme in manchen Ausführungsformen ein oder mehrere Sprachmodelle (zum Beispiel ein großes Sprachmodell), um Ausgaben zu erzeugen, die natürlicher, konversationeller, robuster, skalierbar und genauer sind. Wie bereits erwähnt können aus dem Stand

der Technik bekannte Systeme zum Beispiel digitale oder Gesprächsassistenten verwenden, die mit Fragen und Antworten vorbesetzt sind, was dazu führt, dass die aus dem Stand der Technik bekannten Systeme sowohl in Bezug auf die Fragen, die gestellt werden können, als auch in Bezug auf die Informationen, die als Antwort bereitgestellt werden, beschränkt sind (zum Beispiel beschränkt auf die angegebenen Fragen und Antworten). Im Gegensatz dazu sind die aktuellen Systeme durch die Verwendung des Sprachmodells bzw. der Sprachmodelle nicht auf vordefinierte Fragen und Antworten beschränkt, sondern können Fragen verschiedener Formen interpretieren und im Gegenzug ungeschriebene Antworten unter Verwendung von Daten aus einer Wissensbasis geben. Ferner müssen die Hersteller bei aus dem Stand der Technik bekannten Systemen für jede Marke, jedes Modell und/oder jedes Jahr eines Fahrzeugs oder einer Maschine einen eigenen Gesprächsassistenten erstellen. Im Gegensatz dazu verwenden die aktuellen Systeme in manchen Ausführungsformen sowohl Informationen aus einer oder mehreren Datenbanken als auch ein oder mehrere Sprachmodelle, um Ausgaben zu erzeugen. Daher können die derzeitigen Systeme für viele verschiedene Marken, Modelle und/oder Jahre von Fahrzeugen oder Maschinen verwendet werden.

KURZBESCHREIBUNG DER ZEICHNUNGEN

[0005] Die vorliegenden Systeme und Verfahren für Dialogsysteme für Automobilsysteme und -anwendungen werden im Folgenden unter Bezugnahme auf die beigefügten Zeichnungsfiguren ausführlich beschrieben, wobei:

Fig. 1 ein Beispiel für die Verwendung von Dialogsystemen für Fahrzeuge und Anwendungen gemäß manchen Ausführungsformen der vorliegenden Offenbarung zeigt;

Fig. 2 ein Beispiel für die Verarbeitung von Audiodaten zur Erzeugung von Textdaten, die Sprache darstellen, gemäß manchen Ausführungsformen der vorliegenden Offenbarung zeigt;

Fig. 3 ein Beispiel für Frage/Antwort-Paare gemäß manchen Ausführungsformen der vorliegenden Offenbarung zeigt;

Fig. 4 ein Beispiel für das Abrufen von Frage-/Antwort-Paaren gemäß manchen Ausführungsformen der vorliegenden Offenbarung zeigt;

Fig. 5 ein Beispiel für das Abrufen von Kontextinformationen gemäß manchen Ausführungsformen der vorliegenden Offenbarung zeigt;

Fig. 6 ein Beispiel für die Erzeugung einer Eingabeaufforderung unter Verwendung einer Frage, eines Frage/Antwort-Paares und von

Kontextinformationen, gemäß manchen Ausführungsformen der vorliegenden Offenbarung zeigt;

Fig. 7 ein Ablaufdiagramm ist, das ein Verfahren zur Verwendung einer Frage und eines entsprechenden Frage/Antwort-Paares zeigt, um Informationen zu bestimmen, die einem Fahrzeug zugeordnet sind, gemäß manchen Ausführungsformen der vorliegenden Offenbarung;

Fig. 8 ein Ablaufdiagramm ist, das ein Verfahren zur Verwendung einer Frage und entsprechender Kontextinformationen zeigt, um Informationen zu bestimmen, die einem Fahrzeug zugeordnet sind, gemäß manchen Ausführungsformen der vorliegenden Offenbarung;

Fig. 9A eine Darstellung eines Beispiels für ein autonomes Fahrzeug gemäß manchen Ausführungsformen der vorliegenden Offenbarung ist;

Fig. 9B ein Beispiel für Kamerapositionen und Sichtfelder für das autonome Fahrzeug aus **Fig. 9A** gemäß manchen Ausführungsformen der vorliegenden Offenbarung ist;

Fig. 9C ein Blockdiagramm einer beispielhaften Systemarchitektur für das beispielhafte autonome Fahrzeug aus **Fig. 9A** gemäß manchen Ausführungsformen der vorliegenden Offenbarung ist;

Fig. 9D ein Systemdiagramm für die Kommunikation zwischen dem/den Cloud-basierten Server(n) und dem beispielhaften autonomen Fahrzeug aus **Fig. 9A** gemäß manchen Ausführungsformen der vorliegenden Offenbarung ist;

Fig. 10 ist ein Blockdiagramm einer beispielhaften Computervorrichtung ist, das zur Verwendung bei der Implementierung mancher Ausführungsformen der vorliegenden Offenbarung geeignet ist; und

Fig. 11 ein Blockdiagramm eines beispielhaften Rechenzentrums ist, das sich für die Implementierung mancher Ausführungsformen der vorliegenden Offenbarung eignet.

AUSFÜHRLICHE BESCHREIBUNG

[0006] Es werden Systeme und Verfahren im Zusammenhang mit Dialogsystemen für Automobilsysteme und -anwendungen offenbart. Obgleich die vorliegende Offenbarung in Bezug auf ein Beispiel für ein autonomes oder halbautonomes Fahrzeug 900 (hier alternativ als „Fahrzeug 900“ oder „Ego-Fahrzeug 900“ bezeichnet, von dem ein Beispiel in Bezug auf die **Fig. 9A-9D** beschrieben ist), ist dies nicht als Einschränkung zu verstehen. Zum Beispiel können die hier beschriebenen Systeme und Verfah-

ren ohne Einschränkung von autonomen Fahrzeugen, nicht-autonomen Fahrzeugen, teilautonomen Fahrzeugen (zum Beispiel, in einem oder mehreren adaptiven Fahrerassistenzsystemen (ADAS)), gelenkten und ungelenkten Robotern oder Roboterplattformen, Lagerfahrzeugen, Geländefahrzeugen, Fahrzeugen, die mit einem oder mehreren Anhängern gekoppelt sind, fliegenden Schiffen, Booten, Shuttles, Notarzteinsetzfahrzeugen, Motorrädern, elektrischen oder motorisierten Fahrrädern, Flugzeugen, Baufahrzeugen, Unterwasserfahrzeugen, bemannten oder unbemannten Drohnen und/oder anderen Fahrzeug- oder Maschinentypen verwendet werden. Obgleich die vorliegende Offenbarung in Bezug auf Dialogsysteme beschrieben wird, soll dies nicht einschränkend sein, und die hier beschriebenen Systeme und Verfahren können in den Bereichen Augmented Reality, Virtual Reality, Mixed Reality, Robotik, Sicherheit und Überwachung, autonome oder halbautonome Maschinenanwendungen und/oder in jedem anderen Technologiebereich verwendet werden, in dem Dialogsysteme verwendet werden können.

[0007] Zum Beispiel können ein oder mehrere Systeme Audiodaten empfangen, die mit einem oder mehreren Mikrofonen eines Fahrzeugs erzeugt werden, wobei die Audiodaten Sprache (zum Beispiel eine Äußerung) von einem Benutzer des Fahrzeugs darstellen. In manchen Beispielen kann die Sprache einer Aufgabe zugeordnet sein, die vom Benutzer angefordert wird, wie zum Beispiel eine Eingabeaufforderung zur Bereitstellung von Informationen im Zusammenhang mit dem Fahrzeug. Das (die) System(e) kann (können) dann die Audiodaten unter Verwendung eines (mehrerer) Sprachverarbeitungsmodells (zum Beispiel eines automatischen Spracherkennungsmodells (ASR), eines Modells für die Umwandlung von Sprache in Text (STT), eines Modells für die Verarbeitung natürlicher Sprache (NLP), eines Diarisierungsmodells usw.) verarbeiten, das derart konfiguriert ist, dass es den Audiodaten zugeordnete Textdaten erzeugt. Bei den Textdaten kann es sich zum Beispiel um ein Transkript (zum Beispiel einen oder mehrere Buchstaben, Wörter, Symbole, Zahlen usw.) handeln, das der Sprache zugeordnet ist, und/oder um einen Hinweis darauf, welche(r) Passagier(e)Benutzer(in) der Sprache zugeordnet ist. Wenn die Sprache zum Beispiel einer Anfrage nach Informationen über das Fahrzeug zugeordnet ist, wie zum Beispiel „Welchen Reifendruck sollte ich für die Vorderreifen verwenden“, dann können die Textdaten das Transkript bzw. die Verschriftlichung der Sprache darstellen. In manchen Beispielen können das/die Sprachverarbeitungsmodell(e) die Textdaten ferner generieren, um zusätzliche Informationen darzustellen, die der Sprache zugeordnet sind, wie zum Beispiel eine Absicht der Sprache (zum Beispiel Reifeninformationen erhalten) und/oder Informationen für einen Steckplatz

(zum Beispiel Reifendruck), der der Absicht zugeordnet ist.

[0008] Das (die) System(e) kann (können) dann zusätzliche Informationen abrufen, die den Audiodaten zugeordnet sind. Zum Beispiel können das/die System(e) zunächst Frage/Antwort-Paare in einer oder mehreren Datenbanken oder Datenspeichern speichern. Wie hier beschrieben, kann ein Frage/Antwort-Paar Textdaten enthalten, die eine Frage und eine entsprechende Antwort darstellen. Zum Beispiel kann ein Frage/Antwort-Paar Textdaten umfassen, die eine Frage wie „Welcher Reifendruck wird empfohlen?“ und eine entsprechende Antwort wie „Der Reifendruck sollte auf sechszunddreißig Pfund pro Quadratzoll eingestellt werden“ darstellen. In manchen Beispielen sind die Frage/Antwort-Paare einem oder mehreren Fahrzeugen zugeordnet, zum Beispiel einer oder mehreren Komponenten des oder der Fahrzeuge, einem oder mehreren Merkmalen des oder der Fahrzeuge, einem oder mehreren Wartungsplänen, die dem oder den Fahrzeugen zugeordnet sind, und/oder dergleichen. In manchen Beispielen sind eine oder mehrere der Frage/Antwort-Paare im Allgemeinen mehreren Fahrzeugtypen zugeordnet. Zum Beispiel können ein oder mehrere Frage/Antwort-Paare Autos, Lieferwagen, Lastwagen, Fahrzeugherstellern, Fahrzeugmodellen und/oder dergleichen zugeordnet sein. In manchen Beispielen können ein oder mehrere Frage/Antwort-Paare einem bestimmten Fahrzeugtyp zugeordnet sein, wie zum Beispiel einem bestimmten Fahrzeughersteller, einem bestimmten Fahrzeugmodell und/oder einem bestimmten Fahrzeugtyp.

[0009] Das (die) System(e) kann (können) dann ein (mehrere) Abfragesystem(e) verwenden, um aus der (den) Datenbank(en) (oder Datenspeichern oder anderen Speicher- oder Speicherarten) ein oder mehrere Frage/Antwort-Paare abzurufen, die mit den Textdaten in Zusammenhang stehen. In manchen Beispielen können die in der/den Datenbank(en) gespeicherten Frage/Antwort-Paare Einbettungen zugeordnet sein, um das/die Frage/Antwort-Paar(e) abzurufen. So kann zum Beispiel ein erstes Frage/Antwort-Paar einer ersten Einbettung zugeordnet werden, ein zweites Frage/Antwort-Paar einer zweiten Einbettung, ein drittes Frage/Antwort-Paar einer dritten Einbettung und/oder so weiter. Das (die) Abfragesystem(e) kann (können) die Textdaten verarbeiten, die das Transkript (zum Beispiel die Frage) darstellen, und auf Grundlage der Verarbeitung eine Einbettung für das Transkript erzeugen. Das (die) Abfragesystem(e) kann (können) dann die erzeugte Einbettung verwenden, um das (die) Frage/Antwort-Paar(e) abzurufen. Zum Beispiel können das oder die Abfragesystem(e) die erzeugte Einbettung und die den Frage/Antwort-Paaren zugeordneten Einbettungen verwenden, um die Punktzahlen für die Frage/Antwort-Paare zu bestimmen. Das (die)

Suchsystem(e) kann (können) dann eine Schwellenanzahl der Frage/Antwort-Paare abrufen, die der (den) höchsten Punktzahl(en) zugeordnet sind. Wie hierin beschrieben, kann die Schwellenanzahl von Frage/Antwort-Paaren ein Frage/Antwort-Paar, fünf Frage/Antwort-Paare, zehn Frage/Antwort-Paare, zwanzig Frage/Antwort-Paare und/oder eine beliebige andere Anzahl von Frage/Antwort-Paaren umfassen, ohne jedoch hierauf beschränkt zu sein.

[0010] Während in diesem Beispiel die Verwendung von Einbettungen zum Abrufen der Frage/Antwort-Paare beschrieben wird, kann das Abfragesystem in anderen Beispielen eine oder mehrere zusätzliche und/oder alternative Techniken verwenden. In einem ersten Beispiel können die Frage/Antwort-Paare in verschiedene Kategorien unterteilt werden. Wenn sich die Frage/Antwort-Paare zum Beispiel auf ein oder mehrere Fahrzeuge beziehen, können die Frage/Antwort-Paare in Bauteilkategorien (zum Beispiel Reifen, Motoren, Türen, Fenster usw.), Ausstattungskategorien (zum Beispiel Funkgeräte, Anzeigen usw.), Wartungskategorien (zum Beispiel Zeiträume für empfohlene Wartungsarbeiten usw.) und/oder jede andere Kategorie unterteilt werden. Das (die) Abfragesystem(e) kann (können) dann die Kategorien verwenden, um ein Frage/Antwort-Paar abzurufen, das einer ähnlichen Kategorie angehört wie das durch die Textdaten dargestellte Transkript. In einem zweiten Beispiel können das oder die Abfragesysteme ein oder mehrere Wörter aus dem durch die Textdaten dargestellten Transkript mit einem oder mehreren Wörtern abgleichen, die durch ein oder mehrere Frage/Antwort-Paare dargestellt werden. Das (die) Abfragesystem(e) kann (können) dann ein (mehrere) Frage/Antwort-Paar(e) abrufen, das (die) zumindest eine Schwellenanzahl (zum Beispiel ein, zwei, drei, fünf, zehn usw.) von übereinstimmenden Wörtern enthält (enthalten).

[0011] In manchen Ausführungsformen kann das System zusätzlich oder alternativ zum Speichern und/oder Abrufen von Frage/Antwort-Paaren Absichten, Teil-Absichten, Tokens oder andere Klassifizierungstypen speichern, die verschiedenen Antworttypen entsprechen. In solchen Beispielen können Einbettungen mit der nächstgelegenen Absicht -(zum Beispiel „Reifendruck-Absicht“, „Tankbehälter-öffnen-Absicht“ usw.) abgeglichen werden, und diese Information kann zur Bestimmung einer Antwort verwendet werden. So können Informationen gespeichert werden, die einer Frage und/oder Antwort entsprechen oder diese darstellen, und nicht das Frage/Antwort-Paar selbst.

[0012] Zusätzlich oder alternativ zur Abfrage des Frage-Antwort-Paares kann das Abfragesystem kontextbezogene Informationen abrufen, die mit Textdaten zugeordnet sind. So können das/die System(e) zum Beispiel Informationen speichern, die einem

oder mehreren Fahrzeugen zugeordnet sind. In manchen Beispielen können die Informationen einem bestimmten Fahrzeugtyp zugeordnet sein, wie etwa einem bestimmten Fahrzeughersteller, einem bestimmten Fahrzeugmodell und/oder einem bestimmten Fahrzeugjahr. Die Informationen können zum Beispiel Text aus einer festen oder lebendigen textbasierten Wissensbasis enthalten - wie zum Beispiel in Fahrzeug- oder Maschinenimplementierungen ein OEM-Handbuch, das dem Hersteller, dem Modell und/oder dem Jahr des Fahrzeugs oder der Maschine zugeordnet ist. In manchen Beispielen können die Informationen in der Wissensbasis im Allgemeinen mehreren Fahrzeugtypen zugeordnet sein. Zum Beispiel können die Informationen Text aus mehreren OEM-Handbüchern enthalten, die mehreren Fahrzeugmodellen zugeordnet sind. In manchen Beispielen können die Informationen jedoch auch aus anderen Quellen als den OEM-Handbüchern stammen, zum Beispiel aus einer oder mehreren Netzwerkressourcen, auf die das/die Abfragesystem(e) zugreifen können.

[0013] Das (die) Abfragesystem(e) kann (können) dann zumindest einen Teil der Informationen abrufen, die mit Textdaten zugeordnet sind. In manchen Beispielen können die in der/den Datenbank(en) gespeicherten Informationen Einbettungen zugeordnet sein, um den Teil der Informationen abzurufen, ähnlich wie bei den obigen Frage/Antwort-Paaren. Zum Beispiel kann, wie hier ausführlicher beschrieben, ein erster Teil der Informationen einer ersten Einbettung zugeordnet werden, ein zweiter Teil der Informationen einer zweiten Einbettung, ein dritter Teil der Informationen einer dritten Einbettung und/oder so weiter. Das (die) Abfragesystem(e) kann (können) die erzeugte Einbettung und die den Informationsteilen zugeordneten Einbettungen verwenden, um Bewertungen für die Informationsteile zu bestimmen. Das (die) Abfragesystem(e) kann (können) dann eine Schwellenmenge der Informationen abrufen, die der (den) höchsten Bewertung(en) zugeordnet ist. In manchen Beispielen kann die Schwellenmenge der Informationen eine Schwellenanzahl von Teilen umfassen, wie zum Beispiel ein Teil der Informationen, zwei Teile der Informationen, fünf Teile der Informationen und/oder eine beliebige andere Anzahl von Teilen der Informationen, ohne darauf beschränkt zu sein. Zusätzlich oder alternativ kann in manchen Beispielen die Schwellenmenge der Information eine Schwellenanzahl von Wörtern umfassen, wie zum Beispiel ein Wort der Information, zehn Wörter der Information, einhundert Wörter der Information, zweihundert Wörter der Information und/oder eine beliebige andere Anzahl von Wörtern der Information.

[0014] Während in diesem Beispiel das Verwenden von Einbettungen zum Abrufen der Informationsabschnitte beschrieben wird, können die Abfragesys-

teme in anderen Beispielen eine oder mehrere zusätzliche und/oder alternative Techniken verwenden. In einem ersten Beispiel können die Teile der Informationen in verschiedene Kategorien unterteilt werden. Wenn die Informationen (aus einer Wissensbasis) zum Beispiel einem OEM-Handbuch eines Fahrzeugs zugeordnet sind, können die Teile der Informationen in Komponentenkategorien (zum Beispiel Reifen, Motoren, Türen, Fenster usw.), Ausstattungskategorien (zum Beispiel Radios, Anzeigen usw.), Wartungskategorien (zum Beispiel Zeiträume für empfohlene Wartungsarbeiten usw.) und/oder jede andere Kategorie unterteilt werden. Das (die) Abfragesystem(e) kann (können) dann die Kategorien verwenden, um einen Teil (Teile) der Informationen abzurufen, der (die) einer ähnlichen Kategorie angehört (angehören) wie das durch die Textdaten dargestellte Transkript. In einem zweiten Beispiel können das bzw. die Abfragesystem(e) ein oder mehrere Wörter aus dem durch die Textdaten dargestellten Transkript mit einem oder mehreren Wörtern abgleichen, die durch einen oder mehrere Teile der Informationen dargestellt werden. Das (die) Abfragesystem(e) kann (können) dann einen Teil (Teile) der Informationen abrufen, der (die) zumindest eine Schwellenzahl (zum Beispiel ein, zwei, drei, fünf, zehn usw.) von übereinstimmenden Wörtern enthält (enthalten).

[0015] Das (die) System(e) kann (können) dann die Textdaten, die das Transkript darstellen, Daten, die das (die) Frage/Antwort-Paar(e) darstellen, Daten, die den (die) Teil(e) der Kontextinformationen darstellen, und/oder zusätzliche Daten verwenden, um eine der Sprache zugeordnete Eingabeaufforderung zu erzeugen. Das (die) System(e) kann (können) dann in ein (mehrere) Sprachmodell(e) (zum Beispiel ein großes Sprachmodell(e)) Eingabeaufforderungsdaten eingeben, welche die Eingabeaufforderung darstellen. Wie hierin beschrieben, kann (können) das (die) Sprachmodell(e) jede Art von Sprachmodell(en) umfassen, wie zum Beispiel ein großes Sprachmodell (LLM), generative(s) Sprachmodell(e) (zum Beispiel ein Generative Pretrained Transformer (GPT) usw.), ein Repräsentationssprachmodell(e) (zum Beispiel ein Bidirectional Encoder Representations from Transformers (BERT) usw.) und/oder jede andere Art von Sprachmodell. Das/die Sprachmodell(e) kann/können dann die Eingabeaufforderungsdaten verarbeiten und auf Grundlage der Verarbeitung der Sprache zugeordnete Daten ausgeben. Wenn die Sprache zum Beispiel eine dem Fahrzeug zugeordnete Frage darstellt, können die Ausgabedaten Informationen (zum Beispiel eine Antwort) darstellen, die der Frage zugeordnet sind. Das (die) System(e) kann (können) dann die Ausgabe für den Benutzer, zum Beispiel durch Ausgabe von Audio, das der Ausgabe zugeordnet ist, über einen oder mehrere Lautsprecher bereitstellen.

[0016] Die hier beschriebenen Systeme und Verfahren können ohne Einschränkung von nicht autonomen Fahrzeugen, teilautonomen Fahrzeugen (zum Beispiel in einem oder mehreren adaptiven Fahrerassistenzsystemen (ADAS)), gelenkten und ungelenkten Robotern oder Roboterplattformen, Lagerfahrzeugen, Geländefahrzeugen, Fahrzeugen, die mit einem oder mehreren Anhängern gekoppelt sind, fliegenden Schiffen, Booten, Shuttles, Notarzteinsatzfahrzeugen, Motorrädern, elektrischen oder motorisierten Fahrrädern, Flugzeugen, Baufahrzeugen, Unterwasserfahrzeugen, Drohnen und/oder anderen Fahrzeugtypen verwendet werden. Darüber hinaus können die hierin beschriebenen Systeme und Verfahren für eine Vielzahl von Zwecken verwendet werden, zum Beispiel und ohne hierauf beschränkt zu sein, für Maschinensteuerung, Maschinenfortbewegung, Maschinenantrieb, Erzeugung synthetischer Daten, Modelltraining, Wahrnehmung, erweiterte Realität, virtuelle Realität, gemischte Realität, Robotik, Sicherheit und Überwachung, Simulation und digitales Twinning, Dialogsysteme, autonome oder halbautonome Maschinenanwendungen, Deep Learning, Umweltsimulation, Objekt- oder Aktuator-Simulation und/oder digitales Twinning, Rechenzentrumsverarbeitung, KI-Gespräche, Simulation von Lichttransporten (zum Beispiel Raytracing, Path Tracing), kollaborative Inhaltserstellung für 3D-Assets, Cloud Computing und/oder andere geeignete Anwendungen.

[0017] Die offenbarten Ausführungsformen können in einer Vielzahl unterschiedlicher Systeme enthalten sein, wie zum Beispiel in Automobilsystemen (zum Beispiel, ein Steuerungssystem für eine autonome oder halbautonome Maschine, ein Wahrnehmungssystem für eine autonome oder halbautonome Maschine), Systeme, die unter Verwendung eines Roboters implementiert sind, Luftfahrtsysteme, medizinische Systeme, Bootssysteme, intelligente Bereichsüberwachungssysteme, Systeme zur Durchführung von Deep-Learning-Operationen, Systeme zur Durchführung von Simulationsoperationen, Systeme zur Durchführung von digitalen Twinning-Operationen, Systeme, die unter Verwendung einer Edge-Vorrichtung implementiert sind, Systeme, die eine oder mehrere virtuelle Maschinen (VMs) enthalten, Systeme zur Durchführung von Operationen zur Erzeugung synthetischer Daten, Systeme, die zumindest teilweise in einem Datenzentrum implementiert sind, Systeme zur Durchführung von KI-Gesprächs-Operationen, Systeme zur Durchführung von Lichttransport-Simulationen, Systeme zur Durchführung kollaborativer Inhaltserstellung für 3D-Assets, Systeme, die zumindest teilweise unter Verwendung von Cloud-Computing-Ressourcen implementiert sind, und/oder andere Arten von Systemen.

[0018] Unter Bezugnahme auf **Fig. 1** ist **Fig. 1** ein Beispiel für die Verwendung von Dialogsystemen für Kraftfahrzeuge oder andere Maschinensysteme und -anwendungen gemäß manchen Ausführungsformen der vorliegenden Offenbarung. Es wird angemerkt, dass diese und andere hier beschriebene Anordnungen nur als Beispiele dienen. Andere Anordnungen und Elemente (zum Beispiel Maschinen, Schnittstellen, Funktionen, Anordnungen, Gruppierungen von Funktionen usw.) können zusätzlich zu den gezeigten oder anstelle von ihnen verwendet werden, und manche Elemente können ganz entfallen. Ferner sind viele der hier beschriebenen Elemente funktionale Einheiten, die als einzelne oder verteilte Komponenten oder in Verbindung mit anderen Komponenten und in jeder geeigneten Kombination und an jedem geeigneten Ort implementiert sein können. Verschiedene hier beschriebene Funktionen, die von Einheiten ausgeführt werden, können von Hardware, Firmware und/oder Software ausgeführt werden. Zum Beispiel können verschiedene Funktionen von einem Prozessor ausgeführt werden, der im Speicher gespeicherte Anweisungen ausführt. In manchen Ausführungsformen können die hierin beschriebenen Systeme, Verfahren und Prozesse unter Verwendung ähnlicher Komponenten, Merkmale und/oder Funktionen wie die des autonomen Beispielfahrzeugs 900 der **Fig. 9A-9D**, der beispielhaften Rechenvorrichtung 1000 aus **Fig. 10** und/oder dem beispielhaften Datenzentrum 900 aus **Fig. 11**.

[0019] Der Prozess 100 kann eine oder mehrere sprachverarbeitende Komponenten 102 umfassen, die Audiodaten 104 verarbeiten. Zum Beispiel kann das Fahrzeug die Audiodaten 104 mit Hilfe eines oder mehrerer Mikrofone erzeugen, wobei die Audiodaten 104 Sprache (zum Beispiel eine Äußerung) von einem Benutzer des Fahrzeugs darstellen. In manchen Beispielen kann die Sprache eine Aufgabe darstellen, die vom Benutzer angefordert wird, zum Beispiel eine Frage zum Fahrzeug. Wie hier beschrieben kann die Frage, wenn die Aufgabe eine Frage beinhaltet, einem Merkmal (zum Beispiel einem Radio, einer Anzeige usw.) des Fahrzeugs, einer Komponente (zum Beispiel einem Fenster, einer Tür, einem Reifen, einem Motor usw.) des Fahrzeugs, einem dem Fahrzeug zugeordneten Wartungsplan und/oder einem anderen Aspekt des Fahrzeugs zugeordnet sein. Das Fahrzeug kann dann die Audiodaten 104 mit Hilfe der Sprachverarbeitungskomponente(n) 102 verarbeiten. Wie hier beschrieben kann die sprachverarbeitende(n) Komponente(n) 102 ein oder mehrere ASR-Modelle, ein oder mehrere STT-Modelle, ein oder mehrere NLP-Modelle und/oder jede andere Art von Sprachmodell umfassen, ohne jedoch darauf beschränkt zu sein.

[0020] In manchen Beispielen können die sprachverarbeitende(n) Komponente(n) 102 auf Grundlage der Verarbeitung Textdaten 106 erzeugen, die ein

oder mehrere Wörter (zum Beispiel ein Transkript) darstellen, die der Sprache zugeordnet sind. Wenn die Audiodaten 104 zum Beispiel eine Frage darstellen, die „Was ist der empfohlene Reifendruck?“ beinhaltet, dann können die Textdaten 106 Text darstellen, der „Was ist der empfohlene Reifendruck?“ beinhaltet. In manchen Beispielen können die sprachverarbeitende(n) Komponente(n) 102 die Audiodaten 104 und/oder die Textdaten 106 weiter verarbeiten, um zusätzliche Informationen zu bestimmen, die der Sprache zugeordnet sind. Zum Beispiel können die sprachverarbeitende(n) Komponente(n) 102 ferner eine Absicht der Sprache und/oder Informationen bestimmen, die einem oder mehreren Schlitzes oder Token zugeordnet sind, die sich auf die Absicht beziehen. Wie hier beschrieben kann eine Absicht das Anfordern von Informationen (zum Beispiel Informationen über eine Komponente, ein Merkmal, einen Wartungsplan usw.), das Planen eines Ereignisses (zum Beispiel das Planen eines Wartungstermins usw.) und/oder dergleichen im Zusammenhang mit dem Fahrzeug umfassen, ohne jedoch darauf beschränkt zu sein. In manchen Ausführungsformen können die Textdaten aus einer oder mehreren Eingaben in eine Benutzerschnittstelle, eine Eingabevorrichtung (zum Beispiel eine physische oder digitale Tastatur) und/oder dergleichen erzeugt werden, zusätzlich oder alternativ zu einer Erzeugung auf Grundlage von Audiodaten.

[0021] **Fig. 2** zeigt zum Beispiel ein Beispiel für die sprachverarbeitende(n) Komponente(n) 102, die Audiodaten 202 verarbeitet (welche die Audiodaten 104 darstellen und/oder enthalten können), um Textdaten 204 zu erzeugen (welche die Textdaten 106 darstellen und/oder enthalten können), gemäß manchen Ausführungsformen der vorliegenden Offenbarung. Wie dargestellt können die Audiodaten 104 zumindest Sprache von einem Benutzer darstellen, wobei die Sprache eine Frage ist, die „Was ist der empfohlene Reifendruck?“ umfasst. Die sprachverarbeitende(n) Komponente(n) 102 kann (können) die Audiodaten 202 verarbeiten, um die Textdaten 204 zu erzeugen. Wie dargestellt können die Textdaten 204 zumindest ein Transkript 206 oder eine Aufzeichnung der Sprache darstellen, wie zum Beispiel „Was ist der empfohlene Reifendruck?“. In manchen Beispielen können die Textdaten 204 ferner eine Absicht 208 darstellen, die der Sprache zugeordnet ist, wie zum Beispiel „Informationsanforderung“. In manchen Beispielen können die Textdaten 204 ferner Informationen für Schlitzes 210(1)-(3) darstellen, die der Absicht 208 zugeordnet sind, wobei die Informationen „empfohlen“ für den ersten Schlitz 210(1), „Reifen“ für den zweiten Schlitz 210(2) und „Druck“ für den dritten Schlitz 210(3) umfassen.

[0022] Unter Bezugnahme auf **Fig. 1** kann der Prozess 100 eine Abrufkomponente 108 umfassen, die Frage/Antwort-Daten 110 erzeugt, die den Textdaten

106 zugeordnet sind. Zum Beispiel kann eine Informationsdatenbank (oder mehrere) 112 eine Anzahl von Frage/Antwort-Paaren speichern. Wie hierin beschrieben kann die Anzahl der Frage/Antwort-Paare ein Frage/Antwort-Paar, einhundert Frage/Antwort-Paare, fünfhundert Frage/Antwort-Paare, eintausend Frage/Antwort-Paare und/oder eine beliebige andere Anzahl von Frage/Antwort-Paaren umfassen, ohne jedoch darauf beschränkt zu sein. In manchen Beispielen können die Frage/Antwort-Paare einem bestimmten Fahrzeugtyp zugeordnet sein, wie etwa einem Fahrzeughersteller, einem Fahrzeugmodell und/oder einem Fahrzeugjahr. Die Frage/Antwort-Paare können zum Beispiel mit Hilfe einer Wissensbasis - wie einem OEM-Handbuch - erzeugt werden, die dem Fahrzeug zugeordnet ist. In manchen Beispielen können die Frage/Antwort-Paare mehr als einem Fahrzeugtyp zugeordnet sein. So können die Frage-Antwort-Paare zum Beispiel allgemeine Fragen und Antworten enthalten, die verschiedenen Fahrzeugherstellern, verschiedenen Fahrzeugmodellen und/oder verschiedenen Fahrzeugjahren zugeordnet sind. In manchen Beispielen können die Frage/Antwort-Paare auch anderen Themen als Fahrzeugen zugeordnet sein.

[0023] Als Beispiel für Frage/Antwort-Paare zeigt **Fig. 3** ein Beispiel für Frage/Antwort-Paare für Fahrzeuge gemäß manchen Ausführungsformen der vorliegenden Offenbarung. Wie dargestellt können die Frage/Antwort-Paare eine Anzahl von Fragen 302 (1)-(N) (auch einzeln als „Frage 302“ oder im Plural als „Fragen 302“ bezeichnet) mit entsprechenden Antworten 304(1)-(N) (auch einzeln als „Antwort 304“ oder im Plural als „Antworten 304“ bezeichnet) umfassen. Während das Beispiel aus **Fig. 3** zeigt, dass jede Frage 302 eine entsprechende Antwort 304 umfasst, kann in anderen Beispielen eine Frage 302 mehr als einer Antwort 304 und/oder eine Antwort 304 mehr als einer Frage 302 zugeordnet sein. Während das Beispiel in **Fig. 3** zeigt, dass die Frage/Antwort-Paare Fahrzeugen zugeordnet sind, können die Frage/Antwort-Paare in anderen Beispielen auch jedem anderen Objekttyp und/oder Thema zugeordnet sein.

[0024] Wie hier beschrieben können die Frage/Antwort-Paare in manchen Beispielen einer oder mehreren Kategorien zugeordnet werden. Wenn die Frage/Antwort-Paare zum Beispiel Fahrzeugen zugeordnet sind, dann können die Frage/Antwort-Paare Fahrzeugkomponenten (zum Beispiel Reifen, Fenster, Türen, Motoren usw.), Fahrzeugmerkmalen (zum Beispiel Radios, Anzeigen usw.), Wartungsplänen im Zusammenhang mit den Fahrzeugen (zum Beispiel wann die Bremsen gewartet werden müssen, wann die Reifen gedreht werden müssen usw.) und/oder Ähnlichem zugeordnet sein. In dem Beispiel von **Fig. 3** umfassen die Frage/Antwort-Paare zum Beispiel zumindest sechs Fragen 302(1)-(6) mit

sechs entsprechenden Antworten 304(1)-(6), die Reifen und/oder Reifenwartung zugeordnet sind. Die Frage/Antwort-Paare umfassen ferner eine Frage 302(7) mit einer entsprechenden Antwort 304 (7), die der allgemeinen Wartung zugeordnet ist, eine Frage 302(8) mit einer entsprechenden Antwort 304 (8), die einer Radiofunktion zugeordnet ist, und eine Frage 302(N) mit einer entsprechenden Antwort 304 (N), die einer Türkomponente zugeordnet ist.

[0025] Unter nochmaliger Bezugnahme auf **Fig. 1** kann die Abrufkomponente 108 eine oder mehrere Techniken verwenden, um aus der/den Informationsdatenbank(en) 112 ein oder mehrere der Frage/Antwort-Paare abzurufen, die den Textdaten 106 zugeordnet sind. In manchen Beispielen wird ein Frage/Antwort-Paar den Textdaten 106 basierend auf dem Frage/Antwort-Paar zugeordnet, das sich auf dasselbe Thema, dieselbe Komponente, dasselbe Merkmal und/oder Ähnliches bezieht wie die Frage, die durch die Textdaten 106 dargestellt wird. Zum Beispiel und unter Verwendung des obigen Beispiels, wenn die Frage nach Informationen im Zusammenhang mit Reifen fragt, können die Frage/Antwort-Paare, die der Frage zugeordnet sind, ebenfalls Fragen im Zusammenhang mit Reifen und/oder Antworten umfassen, die Informationen im Zusammenhang mit Reifen umfassen. In manchen Beispielen kann die Abrufkomponente 108 derart konfiguriert sein, dass sie eine Schwellenanzahl von Frage/Antwort-Paaren abrufen. Wie hierin beschrieben kann die Schwellenanzahl von Frage/Antwort-Paaren ein Frage/Antwort-Paar, fünf Frage/Antwort-Paare, zehn Frage/Antwort-Paare, zwanzig Frage/Antwort-Paare und/oder eine beliebige andere Anzahl von Frage/Antwort-Paaren umfassen, ohne jedoch darauf beschränkt zu sein.

[0026] **Fig. 4** zeigt zum Beispiel ein Beispiel für das Abrufen von Frage/Antwort-Paaren gemäß manchen Ausführungsformen der vorliegenden Offenbarung. In dem Beispiel von **Fig. 4** kann die Abrufkomponente 108 eine Art von Abruftechnik verwenden, um die Frage/Antwort-Paare abzurufen, wie zum Beispiel, ohne jedoch darauf beschränkt zu sein, den einbettungsbasierten Abruf. Wie dargestellt sind die Frage/Antwort-Paare (zum Beispiel aus dem Beispiel von **Fig. 3**) Einbettungen 402(1)-(6) zugeordnet (auch einzeln als „Einbettung 402“ oder im Plural als „Einbettungen 402“ bezeichnet). Zum Beispiel ist das Frage/Antwort-Paar, das die Frage 302(1) und die Antwort 304(1) umfasst, der Einbettung 402 (1) zugeordnet, das Frage/Antwort-Paar, das die Frage 302(2) und die Antwort 304(2) enthält, ist der Einbettung 402(2) zugeordnet, das Frage/Antwort-Paar, das die Frage 302(3) und die Antwort 304(3) enthält, ist der Einbettung 402(3) zugeordnet, und/oder so weiter.

[0027] In manchen Beispielen kann die Abrufkomponente 108 die den Frage/Antwort-Paaren zugeordneten Einbettungen 402 erzeugen. So kann die Abrufkomponente 108 zum Beispiel einen oder mehrere Kodierer aufweisen, welche die Frage/Antwort-Paare in dichte Vektoren umwandeln, wobei die Einbettungen 402 den dichten Vektoren zugeordnet sind. In manchen Beispielen können eine oder mehrere andere Komponenten und/oder Systeme die den Frage/Antwort-Paaren zugeordneten Einbettungen 402 erzeugen. Zum Beispiel können die eine oder mehreren anderen Komponenten und/oder Systeme den/die Kodierer aufweisen, der/die die Frage/Antwort-Paare in die dichten Vektoren umwandelt/umwandeln. In manchen Beispielen werden eine oder mehrere der Einbettungen 402 auf Grundlage des Empfangs der Textdaten 204, die das Transkript 206 darstellen, erzeugt. In manchen Beispielen werden eine oder mehrere der Einbettungen 402 vor dem Empfang der Textdaten 204, welche das Transkript 206 darstellen, erzeugt.

[0028] Wie in dem Beispiel aus **Fig. 4** ferner dargestellt, kann die Abrufkomponente 108 ferner eine Einbettung 404 empfangen und/oder erzeugen, die den Textdaten 204 zugeordnet ist. Wie hier beschrieben kann die Abrufkomponente 108 zum Beispiel einen oder mehrere Kodierer aufweisen, welche die Textdaten 204 in einen dichten Vektor umwandeln, wobei die Einbettung 404 dem dichten Vektor zugeordnet ist. Die Abrufkomponente 108 kann dann die Einbettung 404, die den Textdaten 204 zugeordnet ist, und die Einbettungen 402, die den Frage/Antwort-Paaren zugeordnet sind, verwenden (zum Beispiel alle Frage/Antwort-Paare aus dem Beispiel aus **Fig. 3**, obgleich aus Gründen der Übersichtlichkeit nur sechs dargestellt sind), um eine Schwellenanzahl der Frage/Antwort-Paare abzurufen. Zum Beispiel kann die Abrufkomponente 108 die Einbettung 404 und die Einbettungen 402 verwenden, um die Frage/Antwort-Paare zu identifizieren, die dem durch die Textdaten 204 dargestellten Transkript 206 am ähnlichsten sind.

[0029] In manchen Beispielen kann die Abrufkomponente 108 zur Identifizierung der Frage/Antwort-Paare, die dem Transkript 206 am ähnlichsten sind, die Punktzahlen 406(1)-(6) (auch einzeln als „Punktzahl 406“ oder im Plural als „Punktzahlen 406“ bezeichnet) für die Frage/Antwort-Paare unter Verwendung der Einbettung 404 und der Einbettungen 402 bestimmen. Zum Beispiel kann die Abrufkomponente 108 die Punktzahl 406(1) für das Frage/Antwort-Paar, das die Frage 302(1) und die Antwort 304(1) enthält, auf Grundlage der Einbettung 402(1) und der Einbettung 404 bestimmen, die Punktzahl 406(2) für das Frage/Antwort-Paar, das die Frage 302(2) und die Antwort 304(2) umfasst, auf Grundlage der Einbettung 402(2) und der Einbettung 404, die Punktzahl 406(3) für das Frage/Antwort-Paar,

das die Frage 302(3) und die Antwort 304(3) umfasst, auf Grundlage der Einbettung 402(3) und der Einbettung 404, und/oder so weiter. Die Abrufkomponente 108 kann dann eine Schwellenanzahl der Frage/Antwort-Paare auswählen, die den höchsten Punktzahlen zugeordnet sind.

[0030] Zum Beispiel und unter Verwendung des Beispiels aus **Fig. 4** kann die Punktzahl 406(3) für das Frage/Antwort-Paar, das die Frage 302(3) und die Antwort 304(3) umfasst, die höchste Punktzahl umfasst, die Punktzahl 406(1) für das Frage/Antwort-Paar, das die Frage 302(1) und die Antwort 304(1) umfasst, kann die zweithöchste Punktzahl umfassen, die Punktzahl 406(2) für das Frage/Antwort-Paar, das die Frage 302(2) und die Antwort 304(2) umfasst, kann die dritthöchste Punktzahl umfassen, die Punktzahl 406(4) für das Frage/Antwort-Paar, das die Frage 302(4) und die Antwort 304(4) umfasst, kann die vierthöchste Punktzahl umfassen, die Punktzahl 405(6) für das Frage/Antwort-Paar, das die Frage 302(6) und die Antwort 304(6) umfasst, kann die fünfhöchste Punktzahl umfassen, und die Punktzahl 406(5) für das Frage/Antwort-Paar, das die Frage 302(5) und die Antwort 304(5) umfasst, kann die sechste höchste Punktzahl umfassen. Wenn die Abrufkomponente 108 derart konfiguriert ist, dass sie die Frage/Antwort-Paare auswählt, die den drei höchsten Punktzahlen zugeordnet sind (zum Beispiel wenn die Schwellenanzahl der Frage/Antwort-Paare drei Frage/Antwort-Paare beträgt), dann kann die Abrufkomponente 108 die Frage/Antwort-Paare auswählen, welche die Fragen 302(1)-(3) und die Antworten 304(1)-(3) umfassen.

[0031] Während das Beispiel in **Fig. 4** eine beispielhafte Technik veranschaulicht, welche die Abrufkomponente 108 zum Abrufen der Frage/Antwort-Paare verwenden kann, kann die Abrufkomponente 108 in anderen Beispielen zusätzliche und/oder alternative Techniken verwenden. In einem ersten Beispiel können, wie hier beschrieben, die Frage/Antwort-Paare in verschiedene Kategorien unterteilt werden. Zum Beispiel können die Frage/Antwort-Paare, wie in **Fig. 3** dargestellt, in Komponentenkategorien unterteilt werden, wie zum Beispiel Reifen, Motor, Türen, Fenster und/oder dergleichen. In solchen Beispielen kann die Abrufkomponente 108 dann die Kategorien verwenden, um ein Frage/Antwort-Paar abzurufen, das sich in einer ähnlichen Kategorie befindet wie die durch die Textdaten dargestellte Frage. In einem zweiten Beispiel kann die Abrufkomponente 108 ein oder mehrere Wörter, die durch die Textdaten 204 dargestellt werden, mit einem oder mehreren Wörtern abgleichen, die durch ein oder mehrere Frage/Antwort-Paare dargestellt werden. Die Abrufkomponente 108 kann dann ein oder mehrere Frage/Antwort-Paare abrufen, die zumindest eine Schwellenanzahl (zum Beispiel ein, zwei, drei, fünf, zehn usw.) übereinstimmender Wörter aufweisen.

[0032] Unter nochmaliger Bezugnahme auf das Beispiel aus **Fig. 1** kann der Prozess 100 umfassen, dass die Abrufkomponente 108 die Frage/Antwort-Daten 110 erzeugt, die das/die abgerufenen Frage-/Antwort-Paar(e) aus der/den Informationsdatenbank(en) 112 darstellen. In manchen Beispielen können die Frage/Antwort-Daten 110 ferner die Relevanz des Frage/Antwort-Paares/der Frage/Antwort-Paare für die Textdaten 106 darstellen. Wenn die Frage-/Antwort-Daten 110 zum Beispiel mehrere Frage/Antwort-Paare darstellen, dann können die Frage/Antwort-Daten 110 anzeigen, dass ein erstes Frage/Antwort-Paar für die Textdaten 106 am relevantesten ist (zum Beispiel die höchste Punktzahl enthält), ein zweites Frage/Antwort-Paar für die Textdaten 106 am zweitrelevantesten ist (zum Beispiel die zweithöchste Punktzahl enthält), ein drittes Frage/Antwort-Paar für die Textdaten 106 am drittrelevantesten ist (zum Beispiel die dritthöchste Punktzahl enthält) und/oder so weiter.

[0033] Der Prozess 100 kann eine Abrufkomponente 114 umfassen, die kontextbezogene Daten 116 erzeugt, die den Textdaten 106 zugeordnet sind. Zum Beispiel können in der/den Informationsdatenbank(en) 112 auch Informationen gespeichert sein, die einem oder mehreren Fahrzeugen zugeordnet sind. In manchen Beispielen können die Informationen einem bestimmten Fahrzeugtyp zugeordnet sein, wie zum Beispiel einem Fahrzeughersteller, einem Fahrzeugmodell und/oder einem Fahrzeugjahr. Die Informationen können zum Beispiel Text aus einer Wissensbasis (zum Beispiel Handbuch, Dokument, Webseite usw.) umfassen, der dem Fahrzeug zugeordnet ist. In manchen Beispielen können die Informationen im Allgemeinen einer Vielzahl von Fahrzeugtypen zugeordnet sein. Zum Beispiel können die Informationen Text aus mehreren OEM-Handbüchern umfassen, die einer Vielzahl von Fahrzeugmodellen zugeordnet sind. In manchen Beispielen können die Informationen ebenfalls Informationen aus anderen Quellen als OEM-Handbüchern umfassen, wie zum Beispiel Informationen aus einer oder mehreren Netzwerkressourcen, auf die die Abrufkomponente 114 zugreifen kann. In manchen Beispielen können die Informationen ebenfalls anderen Themen als Fahrzeugen zugeordnet sein.

[0034] Die Abrufkomponente 114 kann eine oder mehrere Techniken verwenden, um aus der/den Informationsdatenbank(en) 112 (oder dem Datenspeicher oder einem anderen Speicher oder Speichertyp) Kontextinformationen abzurufen, die den Textdaten 106 zugeordnet sind. In manchen Beispielen werden die Kontextinformationen den Textdaten 106 basierend auf den Kontextinformationen zugeordnet, die sich auf dasselbe Thema, dieselbe Komponente, dasselbe Merkmal und/oder Ähnliches beziehen wie die Frage, die durch die Textdaten 106 dargestellt wird. Wenn die Frage zum Beispiel

nach Informationen im Zusammenhang mit Reifen fragt, dann können die Kontextinformationen, die der Frage zugeordnet sind, Informationen im Zusammenhang mit Reifen enthalten. In manchen Beispielen kann die Abrufkomponente 114 derart konfiguriert sein, dass sie eine Schwellenmenge der kontextbezogenen Informationen abrufen. Wie hier beschrieben kann der Schwellenwert ein Wort, zehn Wörter, hundert Wörter, zweihundert Wörter und/oder jede andere Menge der Kontextinformationen umfassen, ohne jedoch darauf beschränkt zu sein.

[0035] Zum Beispiel veranschaulicht **Fig. 5** ein Beispiel für den Abruf von Kontextinformationen gemäß manchen Ausführungsformen der vorliegenden Offenbarung. In dem Beispiel aus **Fig. 5** sind die kontextbezogenen Informationen einem OEM-Handbuch 502 des Fahrzeugs zugeordnet, in anderen Beispielen können die kontextbezogenen Informationen jedoch jeder anderen Wissensbasis oder Informationsquelle zugeordnet sein. Darüber hinaus kann die Abrufkomponente 114 in dem Beispiel aus **Fig. 5** eine Art des Abrufs verwenden, um die Kontextinformationen abzurufen, wie zum Beispiel den einbettungsbasierten Abruf, ohne darauf beschränkt zu sein.

[0036] Zum Beispiel sind die Informationen 504 aus dem OEM-Handbuch 502, wie dargestellt, in Abschnitte 506(1)-(6) unterteilt (auch einzeln als „Abschnitt 506“ oder im Plural als „Abschnitte 506“ bezeichnet). Jeder der Abschnitte 506 kann eine bestimmte Anzahl von Wörtern enthalten. Während das Beispiel in **Fig. 5** zeigt, dass jeder Abschnitt 506 zehn Wörter enthält, kann in anderen Beispielen jeder Abschnitt 506 eine beliebige andere Anzahl von Wörtern enthalten (zum Beispiel ein Wort, zehn Wörter, fünfzig Wörter, hundert Wörter, zweihundert Wörter usw.). Ferner beginnt jeder Abschnitt 506 mit einer bestimmten Anzahl von Wörtern aus dem vorstehenden Abschnitt 506, zum Beispiel durch Anwendung eines „rollenden“ Verfahrens. In dem Beispiel aus **Fig. 5** beginnt jeder Abschnitt 506 fünf Wörter nach dem vorherigen Abschnitt 506. In anderen Beispielen kann jedoch jeder Abschnitt 506 mit einer beliebigen Anzahl von Wörtern nach dem vorherigen Abschnitt beginnen (zum Beispiel ein Wort, zehn Wörter, fünfzig Wörter, hundert Wörter, zweihundert Wörter usw.).

[0037] Obgleich das Beispiel in **Fig. 5** die Erzeugung der Abschnitte 506 auf Grundlage der Anzahl der Wörter veranschaulicht, können die Abschnitte 506 in anderen Beispielen mit einer oder mehreren zusätzlichen und/oder alternativen Techniken erzeugt werden. So kann in manchen Beispielen jeder Abschnitt 506 einem Teil des OEM-Handbuchs 502 zugeordnet sein. Zum Beispiel kann ein Abschnitt 506 einem Reifenteil des OEM-Handbuchs 502 zugeordnet sein, ein Abschnitt 506 kann einem

Fensterteil des OEM-Handbuchs 502 zugeordnet sein, ein Abschnitt 506 kann einem Motorteil des OEM-Handbuchs 502 zugeordnet sein und/oder so weiter.

[0038] In manchen Beispielen kann die Abrufkomponente 114 derart konfiguriert sein, dass sie die Teile 506 für die Informationen 504 erzeugt. Zum Beispiel kann die Abrufkomponente 114 die Informationen 504 aus dem Handbuch 502 analysieren und auf Grundlage der Analyse die Teile 506 erzeugen. In manchen Beispielen können eine oder mehrere andere Komponenten und/oder Systeme die Teile 506 für die Informationen 504 erzeugen und dann speichern.

[0039] Wie in dem Beispiel aus **Fig. 5** ferner dargestellt sind die Abschnitte 506 den Einbettungen 508 (1)-(6) zugeordnet (auch einzeln als „Einbettung 508“ oder im Plural als „Einbettungen 508“ bezeichnet). Zum Beispiel ist der Teil 506(1) der Einbettung 508 (1), der Teil 506(2) der Einbettung 508(2), der Teil 506(3) der Einbettung 508(3) und/oder so weiter zugeordnet. In manchen Beispielen kann die Abrufkomponente 114 die den Abschnitten 506 zugeordneten Einbettungen 508 erzeugen. Zum Beispiel kann die Abrufkomponente 114 einen oder mehrere Kodierer aufweisen, der die Wörter aus den Abschnitten 506 in dichte Vektoren umwandelt, wobei die Einbettungen 508 den dichten Vektoren zugeordnet sind. In manchen Beispielen können ein oder mehrere andere Systeme und/oder Komponenten die den Abschnitten 506 zugeordneten Einbettungen 508 erzeugen. Zum Beispiel können das eine oder die mehreren anderen Systeme und/oder Komponenten den/die Kodierer aufweisen, der/die die Teile 506 in die dichten Vektoren transformiert/umwandeln. In manchen Beispielen werden eine oder mehrere der Einbettungen 508 auf Grundlage des Empfangs der Textdaten 204, die das Transkript 206 darstellen, erzeugt. In manchen Beispielen werden eine oder mehrere der Einbettungen 508 vor dem Empfang der das Transkript 206 darstellenden Textdaten 204 erzeugt.

[0040] Die Abrufkomponente 114 kann dann die den Textdaten 204 zugeordnete Einbettung 404 und die Einbettungen 508 verwenden, um einen oder mehrere Teile 506 der Informationen 504 auszuwählen. In manchen Beispielen wählt die Abrufkomponente 114 eine Schwellenanzahl der Teile 506 aus. Die Schwellenwertanzahl kann einen Teil 506, zwei Teile 506, fünf Teile 506, zehn Teile 506 und/oder eine beliebige andere Anzahl von Teilen 506 aufweisen, ohne jedoch darauf beschränkt zu sein. In manchen Beispielen wählt die Abrufkomponente 114 die Teile 506 aus, bis eine Schwellenanzahl von Zeichen und/oder Wörtern erreicht ist. Der Schwellenwert für die Anzahl der Wörter kann zum Beispiel zehn Wörter, fünfzig Wörter, hundert Wörter, zweihundert Wör-

ter und/oder eine beliebige andere Anzahl von Wörtern umfassen, ohne jedoch darauf beschränkt zu sein.

[0041] Die Abrufkomponente 114 kann die Einbettung 404 und die Einbettungen 508 verwenden, um einen oder mehrere Abschnitte 506 der Informationen 504 auszuwählen, die dem Transkript 206 der Textdaten 204 am ähnlichsten sind. In manchen Beispielen kann die Abrufkomponente 114 zur Auswahl des/der Abschnitts/e 506, der/die dem Transkript 206 am ähnlichsten ist/sind, Punktzahlen 510(1)-(6) (auch einzeln als „Punktzahl 510“ oder im Plural als „Punktzahlen 510“ bezeichnet) für die Abschnitte 506 unter Verwendung der Einbettung 508 und der Einbettung 404 bestimmen. Zum Beispiel kann die Abrufkomponente 114 die Punktzahl 510(1) für den Abschnitt 506(1) auf Grundlage der Einbettung 508 (1) und der Einbettung 404, die Punktzahl 510(2) für den Abschnitt 506(2) auf Grundlage der Einbettung 508(2) und der Einbettung 404, die Punktzahl 510(3) für die Einbettung 508(3) auf Grundlage der Einbettung 508(3) und der Einbettung 404 und/oder so weiter bestimmen. Die Abrufkomponente 114 kann dann den/die Abschnitt(e) 506 auf Grundlage der Punktzahlen 510 auswählen.

[0042] In Beispielen, in denen die Abrufkomponente 114 nur einen der Teile 506 auswählt, kann die Abrufkomponente 114 den Teil 506 auswählen, der der höchsten Punktzahl 510 zugeordnet ist. Wenn die Abrufkomponente 114 zum Beispiel bestimmt, dass die Punktzahl 510(6) die höchste Punktzahl 510 umfasst, kann die Abrufkomponente 114 den Teil 506(6) der Informationen 504 auswählen. Darüber hinaus kann die Abrufkomponente 114 in Beispielen, in denen die Abrufkomponente 114 mehr als einen der Abschnitte 506 auswählt, wie zum Beispiel eine bestimmte Anzahl von Abschnitten 506, die bestimmte Anzahl von Abschnitten 506 auswählen, welche die höchsten Punktzahlen 510 enthalten. Wenn die Abrufkomponente 114 zum Beispiel zwei Abschnitte 506 auswählt, kann die Abrufkomponente 114 die Abschnitte 506(5)-(6) auf Grundlage der Bestimmung auswählen, dass die Bewertungen 510 (5)-(6) die beiden höchsten Bewertungen 510 enthalten.

[0043] Obgleich das Beispiel in **Fig. 5** eine beispielhafte Technik veranschaulicht, welche die Abrufkomponente 114 zum Abrufen der kontextbezogenen Informationen verwenden kann, kann die Abrufkomponente 114 in anderen Beispielen zusätzliche und/oder alternative Techniken verwenden. In einem ersten Beispiel können die Informationen, wie hier beschrieben, in verschiedene Kategorien unterteilt werden. Zum Beispiel und unter Verwendung des Beispiels von **Fig. 5**, in dem die Informationen 504 aus dem OEM-Handbuch 502 stammen, können die Informationen 504 in Komponentenkate-

gorien wie Reifen, Motor, Türen, Fenster und/oder dergleichen aufgeteilt werden. In solchen Beispielen kann die Abrufkomponente 114 die Kategorien verwenden, um kontextbezogene Informationen abzurufen, die einer ähnlichen Kategorie angehören wie das durch die Textdaten 204 dargestellte Transkript 206. In einem zweiten Beispiel kann die Abrufkomponente 114 ein oder mehrere Wörter, die durch die Textdaten 204 dargestellt werden, mit einem oder mehreren Wörtern abgleichen, die durch die Informationen 504 dargestellt werden. Die Abrufkomponente 114 kann dann kontextbezogene Informationen abrufen, die zumindest eine Schwellenzahl (zum Beispiel ein, zwei, drei, fünf, zehn usw.) übereinstimmender Wörter enthalten.

[0044] Unter nochmaliger Bezugnahme auf das Beispiel aus **Fig. 1** kann der Prozess 100 eine Eingabeaufforderungskomponente 118 aufweisen, welche die Textdaten 106, die Frage/Antwort-Daten 110, die Kontextdaten 116 und/oder zusätzliche Kontextdaten 120 verwendet, um Eingabeaufforderungsdaten 122 zu erzeugen, die den Audiodaten 104 zugeordnet sind. Zum Beispiel kann die Eingabeaufforderungskomponente 118 eine Eingabeaufforderung bzw. einen Prompt erzeugen, indem sie zumindest die durch die Textdaten 106 dargestellte Frage, das/die abgerufenen Frage/Antwort-Paar(e), die durch die Frage/Antwort-Daten 110 dargestellt werden, und die abgerufenen Kontextinformationen, die durch die Kontextdaten 116 dargestellt werden, verwendet. In manchen Beispielen kann die Eingabeaufforderungskomponente 118 derart konfiguriert sein, dass sie die Eingabeaufforderung bzw. Eingabeaufforderung erzeugt, indem sie den Text aus den Textdaten 106, den Text aus den Frage/Antwort-Daten 110 und den Text aus den Kontextdaten 116 in einer bestimmten Reihenfolge anordnet.

[0045] **Fig. 6** zeigt ein Beispiel für die Erzeugung einer Eingabeaufforderung 602 unter Verwendung der Textdaten 106, der Frage-/Antwortdaten 110 und der Kontextdaten 116 gemäß manchen Ausführungsformen der vorliegenden Offenbarung. Wie dargestellt kann die Eingabeaufforderung 602 mit den Kontextinformationen beginnen, die dem Teil 506(6) der Informationen 504 zugeordnet sind, der von der Abrufkomponente 114 abgerufen wurde. Die Eingabeaufforderung 602 kann dann Frage/Antwort-Paare 604(1)-(M) (auch einzeln als „Frage/Antwort-Paar 604“ oder im Plural als „Frage/Antwort-Paare 604“ bezeichnet) umfassen (welche die Frage/Antwort-Paare aus dem Beispiel von **Fig. 3** darstellen können), die von der Abrufkomponente 108 abgerufen wurden. In manchen Beispielen werden die Frage/Antwort-Paare 604 auf Grundlage der Relevanz für die Textdaten 204 angeordnet. In einem ersten Beispiel können die Frage/Antwort-Paare 604 derart angeordnet werden, dass das Frage/Antwort-Paar 604 mit der höchsten Punktzahl an

erster Stelle steht, gefolgt von dem Frage/Antwort-Paar 604 mit der zweithöchsten Punktzahl, gefolgt von dem Frage/Antwort-Paar 604 mit der dritthöchsten Punktzahl, und/oder so weiter. In einem zweiten Beispiel können die Frage/Antwort-Paare 604 derart angeordnet sein, dass das Frage/Antwort-Paar 604 mit der niedrigsten Punktzahl zuerst steht, gefolgt von dem Frage/Antwort-Paar 604 mit der zweitniedrigsten Punktzahl, gefolgt von dem Frage/Antwort-Paar 604 mit der drittniedrigsten Punktzahl, und/oder so weiter. Die Eingabeaufforderung 602 kann dann das den Textdaten 204 (zum Beispiel der Frage) zugeordnete Transkript 206 umfassen. Dies ist nur ein Beispiel für die Anordnung der in der Eingabeaufforderung 602 enthaltenen Informationen; in anderen Beispielen kann die Eingabeaufforderung 602 jede andere Anordnung der Informationen umfassen.

[0046] Darüber hinaus kann die Eingabeaufforderung 602 in manchen Beispielen zusätzliche Zeichen, Symbole und/oder Wörter enthalten, welche die verschiedenen Arten von Informationen voneinander trennen. So kann die Eingabeaufforderung 602 zum Beispiel (1) mit dem Wort „Kontext“ beginnen, (2) gefolgt von einem Doppelpunkt, (3) gefolgt von den Kontextinformationen, (4) gefolgt von zwei Zeilen, (5) gefolgt von dem Wort „Frage“, (6) gefolgt von einem Doppelpunkt, (7) gefolgt von einer Frage, (8) gefolgt von einer Zeile, (9) gefolgt von einer Antwort, (10) gefolgt von einem Doppelpunkt, (11) wobei sich (4)-(10) für jedes Frage/Antwort-Paar wiederholen, (12) gefolgt von dem Transkript. Dies ist zwar nur ein Beispiel für die Verwendung von zusätzlichen Zeichen, Symbolen und/oder Wörtern zur Trennung der verschiedenen Arten von Informationen, aber in anderen Beispielen kann die Eingabeaufforderungskomponente 118 eine oder mehrere zusätzliche und/oder alternative Techniken verwenden.

[0047] Unter nochmaliger Bezugnahme auf das Beispiel aus **Fig. 1** kann der Prozess 100 die Eingabe der Eingabeaufforderungsdaten 122 in ein oder mehrere Sprachmodelle 124 umfassen. Wie hier beschrieben kann das/die Sprachmodell(e) 124 jede Art von Sprachmodell(en) aufweisen, wie zum Beispiel, ohne jedoch darauf beschränkt zu sein, ein generatives Sprachmodell(e) (zum Beispiel ein GPT(s) usw.), ein Repräsentationssprachmodell(e) (zum Beispiel ein BERT(s) usw.) und/oder jede andere Art von Sprachmodell. Das (die) Sprachmodell(e) 124 kann (können) derart konfiguriert sein, dass es (sie) die Eingabeaufforderungsdaten 122 verarbeitet (verarbeiten), und auf Grundlage der Verarbeitung kann (können) das (die) Sprachmodell(e) 124 Daten 126 ausgeben, die den Audiodaten 104 (zum Beispiel mit der Frage) zugeordnet sind. Wenn die Audiodaten 104 zum Beispiel Sprache darstellen, die eine Frage zu einer Komponente des Fahrzeugs enthält, dann können die Ausgabedaten

126 Informationen über die Komponente des Fahrzeugs darstellen.

[0048] Wie ferner dargestellt können das/die Sprachmodell(e) 124 ferner kontextbezogene Daten 120 ausgeben, die in manchen Beispielen zumindest einen Teil der Ausgabedaten 126 enthalten können. Wie hier erläutert kann die Eingabeaufforderungskomponente 118 ferner zumindest einen Teil der Kontextdaten 120 verwenden, um die Eingabeaufforderungsdaten 122 zu erzeugen. Wenn der Benutzer zum Beispiel ferner Fragen im Zusammenhang mit dem Fahrzeug stellt, kann die Eingabeaufforderungskomponente 118 die Kontextdaten 120 verwenden, um ferner die Eingabeaufforderungsdaten 122 für die Fragen zu erzeugen, wobei die Kontextdaten 120 einen Kontext darstellen, der den Ausgaben zu früheren Fragen zugeordnet ist.

[0049] In manchen Beispielen können eine oder mehrere Techniken verwendet werden, um zu bestimmen, ob das dem Prozess 100 zugeordnete Dialogsystem korrekt ist. So kann zum Beispiel eine anfängliche Menge von Frage/Antwort-Paaren erzeugt werden, wobei ein erster Teil der anfänglichen Menge von Frage/Antwort-Paaren Referenz-Frage/Antwort-Paaren und ein zweiter Teil der anfänglichen Menge von Frage/Antwort-Paaren Test-Frage/Antwort-Paaren zugeordnet ist. Wenn zum Beispiel der anfängliche Satz von Frage/Antwort-Paaren fünfhundert Frage/Antwort-Paare umfasst, dann können vierhundert der Frage/Antwort-Paare Referenzfrage/Antwort-Paaren und hundert der Frage/Antwort-Paare Test- oder Validierungsfrage/Antwort-Paaren zugeordnet sein. Die Referenzfrage-/Antwortpaare können dann in der/den Informationsdatenbank(en) 112 gespeichert und zur Durchführung der hier beschriebenen Prozesse verwendet werden. Darüber hinaus können die Testfrage-/Antwortpaare zum Testen des Systems/der Systeme verwendet werden.

[0050] Zum Beispiel kann das System bzw. können die Systeme den Prozess 100 aus **Fig. 1** unter Verwendung einer Frage aus einem Testfrage/Antwort-Paar durchführen (zum Beispiel können die Textdaten 106 die Frage darstellen). Auf Grundlage der Durchführung des Prozesses 100 kann/können das/-die Sprachmodell(e) 124 Ausgabedaten 126 erzeugen, die der Frage zugeordnet sind. Das (die) System(e) kann (können) dann die Antwort aus dem Testfrage/Antwort-Paar mit der durch die Ausgabedaten 126 dargestellten Antwort vergleichen. Darüber hinaus können das (die) System(e) einen ähnlichen Prozess für eines oder mehrere der anderen Testfrage/Antwort-Paare durchführen. Auf Grundlage des Vergleichs bzw. der Vergleiche kann das System bzw. können die Systeme eine dem Dialogsystem zugeordnete Genauigkeit bestimmen. Darüber hinaus kann das System (können die Systeme)

einen oder mehrere Prozesse auf Grundlage der Genauigkeit durchführen.

[0051] Wenn das (die) System(e) zum Beispiel bestimmt (bestimmen), dass das Dialogsystem genau ist, dann kann (können) das (die) System(e) bestimmen, dass die in der (den) Informationsdatenbank(en) 112 gespeicherten Informationen (zum Beispiel die Referenzfrage/Antwort-Paare, die zur Gewinnung der Kontextinformationen verwendeten Informationen usw.) und/oder die zur Erzeugung der Eingabeaufforderungsdaten 122 verwendete(n) Technik(en) das Dialogsystem als genau erscheinen lassen. Wenn das System/die Systeme jedoch bestimmt/bestimmen, dass das Dialogsystem ungenau ist, kann/können das System/die Systeme die in der/den Informationsdatenbank(en) 112 gespeicherten Informationen (zum Beispiel die Referenzfrage-/Antwort-Paare, die zur Gewinnung der Kontextinformationen verwendeten Informationen usw.) und/oder die zur Erzeugung der Eingabeaufforderungsdaten 122 verwendete(n) Technik(en) aktualisieren. In manchen Beispielen können das/die System(e) solche Aktualisierungen vornehmen, um das Dialogsystem genauer zu machen.

[0052] Obgleich das Beispiel in **Fig. 1** zeigt, dass die Abrufkomponente 108 von der Abrufkomponente 114 getrennt ist, können in anderen Beispielen die Abrufkomponente 108 und die Abrufkomponente 114 eine einzige Komponente umfassen, welche die hier beschriebenen Prozesse der Abrufkomponenten 108, 114 durchführt. Darüber hinaus können die Abrufkomponenten 108, 114 jede Art von Komponente(n), System(en), Anwendung(en) und/oder dergleichen umfassen, die derart konfiguriert sind, dass sie nach Informationen aus der/den Informationsdatenbank(en) 112 suchen und/oder diese abrufen.

[0053] Unter Bezugnahme auf **Fig. 7** bis 8 umfasst jeder Block der hier beschriebenen Verfahren 700 und 800 einen Rechenprozess, der der einer beliebigen Kombination aus Hardware, Firmware und/oder Software durchgeführt werden kann. Zum Beispiel können verschiedene Funktionen von einem Prozessor ausgeführt werden, der im Speicher gespeicherte Anweisungen ausführt. Die Verfahren 700 und 800 können auch als computerverwendbare Anweisungen auf Computerspeichermedien gespeichert sein. Die Verfahren 700 und 800 können durch eine eigenständige Anwendung, einen Dienst oder einen gehosteten Dienst (eigenständig oder in Kombination mit einem anderen gehosteten Dienst) oder ein Plug-in für ein anderes Produkt bereitgestellt werden, um nur manche zu nennen. Darüber hinaus werden die Verfahren 700 und 800 beispielhaft in Bezug auf das System aus **Fig. 1** beschrieben. Diese Verfahren 700 und 800 können jedoch zusätzlich oder alternativ von einem beliebigen System oder einer beliebigen Kombination von Systemen ausgeführt werden, ein-

schließlich, aber nicht beschränkt auf die hier beschriebenen Systeme.

[0054] Fig. 7 ist ein Ablaufdiagramm, das ein Verfahren 700 zur Verwendung einer Frage und eines entsprechenden Frage/Antwort-Paares zeigt, um Informationen zu bestimmen, die einem Fahrzeug zugeordnet sind, gemäß manchen Ausführungsformen der vorliegenden Offenbarung. Das Verfahren 700 kann im Block B702 das Erhalten von Textdaten umfassen, die eine einem Fahrzeug zugeordnete Frage darstellen. Zum Beispiel kann das Fahrzeug ein oder mehrere Mikrofone verwenden, um Audiodaten 104 zu erzeugen, die Sprache von einem Fahrzeuginsassen darstellen. Das Fahrzeug kann dann die Audiodaten 104 unter Verwendung der Sprachverarbeitungskomponente(n) 102 verarbeiten, um die Textdaten 106 zu erzeugen, welche die Sprache darstellen. In manchen Beispielen kann die Sprache eine dem Fahrzeug zugeordnete Frage umfassen. So kann die Sprache zum Beispiel eine Frage zu einer Komponente des Fahrzeugs, einem Merkmal des Fahrzeugs, einer Wartung des Fahrzeugs und/oder dergleichen umfassen.

[0055] Das Verfahren 700 kann in Block B704 die Bestimmung eines oder mehrerer Frage/Antwort-Paare, die der Frage zugeordnet sind, zumindest auf Grundlage der Textdaten umfassen. Zum Beispiel kann das Fahrzeug die Abrufkomponente 108 verwenden, um das/die Frage/Antwort-Paar(e) abzurufen, die der Frage zugeordnet sind. Wie hierin beschrieben kann die Abrufkomponente 108 in manchen Beispielen Einbettungen verwenden, die den in der/den Informationsdatenbank(en) 112 gespeicherten Frage/Antwort-Paaren und einer der Frage zugeordneten Einbettung zugeordnet sind, um das/die Frage/Antwort-Paar(e) abzurufen, die für die Frage relevant sind. In anderen Beispielen kann die Abrufkomponente 108 jedoch eine oder mehrere zusätzliche und/oder alternative Techniken verwenden, um das/die Frage/Antwort-Paar(e) aus der/den Informationsdatenbank(en) 112 abzurufen. Die Abrufkomponente 108 kann dann Frage/Antwort-Daten 110 erzeugen, die das/die abgerufenen Frage/Antwort-Paar(e) darstellen.

[0056] Der Prozess 700 kann in Block B706 die Eingabe der Textdaten und der Daten, die ein oder mehrere Frage/Antwort-Paare darstellen, in ein Sprachmodell umfassen. Zum Beispiel kann das Fahrzeug die Textdaten 106 und die Frage/Antwort-Daten 110 in das/die Sprachmodell(e) 124 eingeben. In manchen Beispielen kann die Eingabeaufforderungskomponente 118 zur Eingabe der Daten zunächst die Textdaten 106 und die Frage/Antwort-Daten 110 verwenden, um eine Eingabeaufforderung zu erzeugen. Die Eingabeaufforderungskomponente 118 kann dann Eingabeaufforderungsdaten 122, welche die Eingabeaufforderung darstellen, in das/die

Sprachmodell(e) 124 eingeben. In manchen Beispielen kann das Fahrzeug zusätzliche Daten in das/die Sprachmodell(e) 124 eingeben, zum Beispiel von der Abrufkomponente 114 erzeugte Kontextdaten 116 und/oder zuvor von dem/den Sprachmodell(en) 124 ausgegebene Kontextdaten 120.

[0057] Das Verfahren 700 kann in Block B708 die Bestimmung einer der Frage zugeordneten Ausgabe unter Verwendung des Sprachmodells umfassen. Zum Beispiel können das/die Sprachmodell(e) 124 die Textdaten 106 und die Frage-/Antwortdaten 110 (zum Beispiel die Eingabeaufforderungsdaten 122) verarbeiten und auf Grundlage der Verarbeitung der Frage zugeordnete Ausgabedaten 126 ausgeben. Wie hierin beschrieben können die Ausgabedaten 126 Informationen darstellen, die der Frage zugeordnet sind. Das Fahrzeug kann dann die Informationen an den Fahrgast weitergeben. In manchen Beispielen stellt das Fahrzeug die Informationen bereit, indem es einen mit den Ausgabedaten 126 verknüpften Ton ausgibt, wobei der Ton ein oder mehrere Wörter enthält, welche die Informationen darstellen. In manchen Beispielen stellt das Fahrzeug die Informationen bereit, indem es einen mit den Ausgabedaten 126 zugeordneten Inhalt anzeigt, wobei der Inhalt ein oder mehrere Wörter umfasst, welche die Informationen darstellen.

[0058] Fig. 8 ist ein Ablaufdiagramm, das ein Verfahren 800 zur Verwendung einer Frage und entsprechender Kontextinformationen zeigt, um Informationen zu bestimmen, die einem Fahrzeug zugeordnet sind, gemäß manchen Ausführungsformen der vorliegenden Offenbarung. Das Verfahren 800 kann in Block B802 das Erhalten von Textdaten umfassen, die eine einem Fahrzeug zugeordnete Frage darstellen. Zum Beispiel kann das Fahrzeug ein oder mehrere Mikrofone verwenden, um Audiodaten 104 zu erzeugen, die Sprache von einem Fahrzeuginsassen darstellen. Das Fahrzeug kann dann die Audiodaten 104 unter Verwendung der Sprachverarbeitungskomponente(n) 102 verarbeiten, um die Textdaten 106 zu erzeugen, welche die Sprache darstellen. In manchen Beispielen kann die Sprache eine dem Fahrzeug zugeordnete Frage umfassen. Zum Beispiel kann die Sprache eine Frage zu einer Komponente des Fahrzeugs, einem Merkmal des Fahrzeugs, der Wartung des Fahrzeugs und/oder dergleichen umfassen.

[0059] Das Verfahren 800 kann in Block B804 die Bestimmung von Kontextinformationen, die der Frage zugeordnet sind, zumindest auf Grundlage der Textdaten umfassen. Zum Beispiel kann das Fahrzeug die Abrufkomponente 114 verwenden, um die der Frage zugeordneten Kontextinformationen abzurufen. Wie hierin beschrieben kann die Abrufkomponente 114 in manchen Beispielen Einbettungen verwenden, die Teilen von Informationen, die in

der/den Informationsdatenbank(en) 112 gespeichert sind, und einer Einbettung, die der Frage zugeordnet ist, zugeordnet sind, um die kontextbezogenen Informationen abzurufen, die für die Frage relevant sind. In anderen Beispielen kann die Abrufkomponente 114 jedoch eine oder mehrere zusätzliche und/oder alternative Techniken verwenden, um die kontextbezogenen Informationen aus der/den Informationsdatenbank(en) 112 abzurufen. Die Abrufkomponente 114 kann dann kontextbezogene Daten 116 erzeugen, welche die abgerufenen kontextbezogenen Informationen darstellen.

[0060] Der Prozess 800 kann in Block B806 die Eingabe der Textdaten und der Daten, welche die Kontextinformationen darstellen, in ein Sprachmodell beinhalten. Zum Beispiel kann das Fahrzeug die Textdaten 106 und die Kontextdaten 116 in das/die Sprachmodell(e) 124 eingeben. In manchen Beispielen kann die Eingabeaufforderungskomponente 118 zur Eingabe der Daten zunächst die Textdaten 106 und die Kontextdaten 116 verwenden, um eine Eingabeaufforderung zu erzeugen. Die Eingabeaufforderungskomponente 118 kann dann Eingabeaufforderungsdaten 122, welche die Eingabeaufforderung darstellen, in das/die Sprachmodell(e) 124 eingeben. In manchen Beispielen kann das Fahrzeug zusätzliche Daten in das/die Sprachmodell(e) 124 eingeben, zum Beispiel Frage-/Antwortdaten 110, die von der Abrufkomponente 108 erzeugt werden, und/oder Kontextdaten 120, die zuvor von dem/den Sprachmodell(en) 124 ausgegeben werden.

[0061] Das Verfahren 800 kann in Block B808 die Bestimmung einer der Frage zugeordneten Ausgabe unter Verwendung des Sprachmodells umfassen. Zum Beispiel können das/die Sprachmodell(e) 124 die Textdaten 106 und die Kontextdaten 116 (zum Beispiel die Eingabeaufforderungsdaten 122) verarbeiten und auf Grundlage der Verarbeitung Daten 126 ausgeben, die der Frage zugeordnet sind. Wie hierin beschrieben können die Ausgabedaten 126 Informationen darstellen, die der Frage zugeordnet sind. Das Fahrzeug kann dann die Informationen dem Fahrgast über eine oder mehrere Komponenten (zum Beispiel Lautsprecher, Displays, Head-up-Displays, Instrumententafeln usw.) des Fahrzeugs (oder einer anderen Maschine) bereitstellen oder mitteilen. In manchen Beispielen stellt das Fahrzeug die Informationen bereit, indem es einen Ton ausgibt, der den Ausgabedaten 126 zugeordnet ist, wobei der Ton ein oder mehrere Wörter enthält, welche die Informationen darstellen. In manchen Beispielen stellt das Fahrzeug die Informationen bereit, indem es die Anzeige von Inhalten veranlasst, die den Ausgabedaten 126 zugeordnet sind, wobei die Inhalte ein oder mehrere Wörter enthalten, welche die Informationen darstellen.

[0062] Die hier beschriebene Frage-/Antwort-Interaktion - zum Beispiel in Bezug auf die Methoden 700 und 800 - kann mit einem digitalen oder persönlichen Assistenten durchgeführt werden. Zum Beispiel kann der Benutzer während eines Gesprächs zwischen einem Benutzer und einem digitalen Assistenten - etwa während des Betriebs eines Fahrzeugs oder einer anderen Maschine - dem digitalen oder persönlichen Assistenten Fragen stellen, um mehr über die Maschine zu erfahren, um zu lernen, wie man Wartungsarbeiten durchführt, bestimmte Funktionen aktiviert/deaktiviert usw., und die Antworten des digitalen oder persönlichen Assistenten können sich - zumindest teilweise - auf die Ausgabedaten 126 aus den Sprachmodellen 124 stützen, wie sie unter Verwendung des Systems 100 erzeugt werden.

BEISPIELHAFTES AUTONOMES FAHRZEUG

[0063] Fig. 9A ist eine Veranschaulichung eines beispielhaften autonomen Fahrzeugs 900 gemäß manchen Ausführungsformen der vorliegenden Offenbarung. Das autonome Fahrzeug 900 (hier alternativ als „Fahrzeug 900“ bezeichnet) kann ohne Einschränkung ein Personenfahrzeug beinhalten, wie etwa ein Auto, einen Lastwagen, einen Bus, ein Rettungsfahrzeug, ein Shuttle, ein elektrisches oder motorisiertes Fahrrad, ein Motorrad, ein Feuerwehrauto, ein Polizeifahrzeug, einen Krankenwagen, ein Boot, ein Baufahrzeug, ein Unterwasserfahrzeug, ein Roboterfahrzeug, eine Drohne, ein Flugzeug, ein an einen Anhänger gekoppeltes Fahrzeug (z.B. einen Sattelschlepper zum Transport von Frachtgut) und/oder eine andere Art von Fahrzeug (das z.B. unbemannt ist und/oder einen oder mehrere Passagiere beherbergt). Autonome Fahrzeuge werden im Allgemeinen im Hinblick auf Automatisierungslevels beschrieben, die von der National Highway Traffic Safety Administration (NHTSA), einer Abteilung des US-Verkehrsministeriums, und der Society of Automotive Engineers (SAE) „Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles“ (z.B. Standard Nr. J3016-2010906, veröffentlicht am 15. Juni 2018, Standard Nr. J3016-201609, veröffentlicht am 30. September 2016, sowie frühere und zukünftige Versionen dieses Standards) definiert sind. Das Fahrzeug 900 kann gemäß einer oder mehreren der Stufen 3-5 der autonomen Fahrstufen funktionstüchtig sein. Das Fahrzeug 900 kann gemäß einer oder mehreren der Stufen 1-5 der autonomen Fahrstufen funktionstüchtig sein. Zum Beispiel kann das Fahrzeug 900 je nach Ausführungsform zu einer Fahrerassistenz (Stufe 1), einer teilweisen Automatisierung (Stufe 2), einer bedingten Automatisierung (Stufe 3), einer hohen Automatisierung (Stufe 4) und/oder einer vollständigen Automatisierung (Stufe 5) in der Lage sein. Der Begriff „autonom“, wie er hierin verwendet wird, kann beliebige und/oder alle Arten von Autonomie für das Fahrzeug 900 oder eine andere

Maschine beinhalten, wie etwa vollständig autonom sein, hochgradig autonom sein, bedingt autonom sein, teilweise autonom sein, unterstützende Autonomie bereitstellen, halbautonom sein, primär autonom sein oder eine andere Bezeichnung haben.

[0064] Das Fahrzeug 900 kann Komponenten wie Chassis, Karosserie, Räder (z.B. 2, 4, 6, 8, 18 usw.), Reifen, Achsen und andere Komponenten eines Fahrzeugs beinhalten. Das Fahrzeug 900 kann ein Antriebssystem 950 enthalten, wie z.B. einen Verbrennungsmotor, eine Hybrid-Elektro-Antriebsmaschine, einen vollelektrischen Motor und/oder einen anderen Antriebstyp. Das Antriebssystem 950 kann an einen Antriebsstrang des Fahrzeugs 900 angeschlossen werden, der ein Getriebe enthalten kann, um den Antrieb des Fahrzeugs 900 zu ermöglichen. Das Antriebssystem 950 kann als Reaktion auf den Empfang von Signalen des Gaspedals/Gashebels 952 gesteuert werden.

[0065] Ein Lenksystem 954, das möglicherweise ein Lenkrad enthält, kann verwendet werden, um das Fahrzeug 900 zu lenken (z.B. entlang eines gewünschten Wegs oder einer gewünschten Route), wenn das Antriebssystem 950 in Betrieb ist (z.B. wenn das Fahrzeug in Bewegung ist). Das Lenksystem 954 kann Signale von einem Lenkaktor 956 empfangen. Für die vollständige Automatisierungsfunktionalität (Stufe 5) kann das Lenkrad optional sein.

[0066] Das Bremssensorsystem 946 kann zur Betätigung der Fahrzeugbremsen verwendet werden, wenn Signale von den Bremszylindern 948 und/oder Bremssensoren empfangen werden.

[0067] Controller 936, die ein oder mehrere System-on-Chips (SoCs) 904 (**Fig. 9C**) und/oder GPU umfassen, können Signale (z.B. repräsentativ für Befehle) an eine oder mehrere Komponenten und/oder Systeme des Fahrzeugs 900 senden. Zum Beispiel kann der Controller Signale senden, um die Fahrzeugbremse über einen oder mehrere Bremszylinder 948 zu betätigen, das Lenksystem 954 über einen oder mehrere Lenkaktoren 956 zu bedienen, um das Antriebssystem 950 über eines oder mehrere Gaspedale/Gashebel 952 zu bedienen. Der Controller 936 kann ein oder mehrere bordeigene (z.B. integrierte) Rechenvorrichtungen (z.B. Supercomputer) umfassen, die Sensorsignale verarbeiten, und Betriebsbefehle (z.B. Signale, die Befehle darstellen) ausgeben, um autonomes Fahren zu ermöglichen und/oder einen menschlichen Fahrer beim Fahren des Fahrzeugs zu unterstützen 900. Der Controller 936 kann einen ersten Controller 936 für autonome Fahrfunktionen, einen zweiten Controller 936 für funktionale Sicherheitsfunktionen, einen dritten Controller 936 für die Funktionalität der künstlichen Intelligenz (z.B. Computervision), einen vierten Controller

936 für Infotainment-Funktionalität, einen fünften Controller 936 für Redundanz im Notfall und/oder andere Controller enthalten. In manchen Beispielen kann ein einzelner Controller 936 zwei oder mehr der oben genannten Funktionalitäten verarbeiten, zwei oder mehr Controller 936 können eine einzelne Funktionalität und/oder eine beliebige Kombination davon verarbeiten.

[0068] Der Controller 936 kann die Signale zur Steuerung einer oder mehrerer Komponenten und/oder Systeme des Fahrzeugs 900 als Reaktion auf Sensordaten liefern, die von einem oder mehreren Sensoren (z.B. Sensoreingaben) empfangen werden. Die Sensordaten können zum Beispiel und ohne Einschränkung von globalen Satellitennavigationssystem-Sensoren 958 (z.B. Global Positioning System Sensor(en)), RADAR-Sensor(en) 960, Ultraschallsensor(en) 962, LIDAR-Sensor(en) 964, Trägheitsmesseinheiten-Sensor(en) (IMU) 966 (z.B. Beschleunigungsmesser, Gyroskop(e), Magnetkompass(e), Magnetometer usw.), Mikrofon(en) 996996, Stereokamera(s) 968, Weitwinkelkamera(s) 970 (z.B. Fischaugenkamera(s), Infrarotkamera(s) 972, Surround-Kamera(s) 974 (z.B. 360-Grad-Kameras), Long-Range-Kamera(s) und/oder Mid-Range-Kameras 998, Geschwindigkeitssensor(en) 944 (z.B. zur Messung der Fahrzeuggeschwindigkeit 900), Vibrationssensor(en) 942, Lenksensor(en) 940, Bremsensor(en) (z.B. als Teil des Bremssensorsystems 946) und/oder anderen Sensortypen empfangen werden.

[0069] Eine oder mehrere der Controller 936 können Eingaben (z.B. dargestellt durch Eingabedaten) von einem Kombiinstrument 932 des Fahrzeugs 900 empfangen und Ausgaben (z.B. dargestellt durch Ausgabedaten, Anzeigedaten usw.) über eine Mensch-Maschine-Schnittstelle (human-machine interface - HMI)-Anzeige 934, einen akustischen Signalgeber, einen Lautsprecher und/oder über andere Komponenten des Fahrzeugs 900 empfangen. Die Ausgaben können Informationen wie Fahrzeuggeschwindigkeit, Geschwindigkeit, Zeit, Kartendaten (z.B. die HD-Karte 922 aus **Fig. 9C**), Positionsdaten (z.B. die Position des Fahrzeugs 900, wie etwa auf einer Karte), Richtung, Position anderer Fahrzeuge (z.B. ein Belegungsrastrer), Informationen über Objekte und den Status von Objekten, wie sie von den Controllern 936 wahrgenommen werden usw. beinhalten. Auf der HMI-Anzeige 934 können zum Beispiel Informationen über das Vorhandensein eines oder mehrerer Objekte (z.B. Straßenschild, Warnschild, Ampelschaltungen usw.) und/oder Informationen über Fahrmanöver angezeigt werden, die das Fahrzeug unternommen hat, gerade macht oder durchführen wird (z.B. jetzt Spurwechsel, Ausfahrt 34B in zwei Meilen nehmen usw.).

[0070] Das Fahrzeug 900 verfügt außerdem über eine Netzwerkschnittstelle 924, die eine oder mehrere Wireless-Antennen 926 und/oder Modem(s) zur Kommunikation über ein oder mehrere Netzwerke verwenden kann. Die Netzwerkschnittstelle 924 kann zum Beispiel über LTE (Long Term Evolution), WCDMA (Wideband Code Division Multiple Access), UMTS (Universal Mobile Telecommunications System), GSM (Global System for Mobile communication), CDMA2000 (IMT-CDMA Multi-Carrier) usw. kommunizieren. Die Wireless-Antenne(n) 926 kann/können auch die Kommunikation zwischen Objekten in der Umgebung (z.B. Fahrzeugen, mobilen Geräten usw.) über lokale Netzwerke wie Bluetooth, Bluetooth Low Energy (LE), Z-Wave, ZigBee usw. und/oder Low-Power-Wide-Area-Network (LPWAN) wie LoRaWAN, SigFox usw. ermöglichen.

[0071] Fig. 9B ist ein Beispiel für Kamerapositionen und Sichtfelder für das beispielhafte autonome Fahrzeug 900 aus Fig. 9A, gemäß manchen Ausführungsformen der vorliegenden Offenbarung. Die Kameras und die entsprechenden Sichtfelder stellen eine beispielhafte Ausführungsform dar und sind nicht als einschränkend aufzufassen. Zum Beispiel können zusätzliche und/oder alternative Kameras enthalten sein und/oder die Kameras können sich an verschiedenen Positionen im Fahrzeug 900 befinden.

[0072] Die Kameratypen für die Kameras können, sind aber nicht darauf beschränkt, unter anderem Digitalkameras umfassen, die für die Verwendung mit den Komponenten und/oder Systemen des Fahrzeugs 900 angepasst werden können. Die Kamera(s) können mit dem Automobilsicherheitsintegritätslevel (automotive safety integrity level - ASIL) B und/oder mit einem anderen ASIL betrieben werden. Die Kameraarten können in Abhängigkeit von der Ausführungsform zu einer beliebigen Bildaufnahme rate in der Lage sein, wie etwa 60 Bilder pro Sekunde (frames per second - fps), 120 fps, 240 fps usw. Die Kameras können dazu in der Lage sein, Rollblendenverschlüsse, globale Blendenverschlüsse, einen anderen Typ von Blendenverschluss oder eine Kombination davon zu verwenden. In manchen Beispielen kann die Farbfilteranordnung eine Rot-Klar-Klar-Klar (red clear clear clear - RCCC)-Farbfilteranordnung, eine Rot-Klar-Klar-Blau (red clear clear blue - RCCB)-Farbfilteranordnung, eine Rot-Blau-Grün-Klar (red blue green clear - RBGC)-Farbfilteranordnung, eine Foveon-X3-Farbfilteranordnung, eine Bayer-Sensoren (RGGB)-Farbfilteranordnung, eine Monochrom-Sensor-Farbfilteranordnung und/oder eine andere Art von Farbfilteranordnung beinhalten. In manchen Ausführungsformen können Klarpixelkameras, wie zum Beispiel Kameras mit einer RCCC-, einer RCCB- und/oder einer RBGC-Farbfilteranordnung, in einem Bestreben zur Erhöhung der Lichtempfindlichkeit verwendet werden.

[0073] In manchen Ausführungsformen können eine oder mehrere der Kamera(s) verwendet werden, um Funktionen der weiterentwickelten Fahrerassistenzsysteme (advanced driver assistance systems - ADAS) durchzuführen (z.B. als Teil einer redundanten oder ausfallsicheren Ausgestaltung). Zum Beispiel kann eine Multifunktions-Monokamera installiert sein, die Funktionen wie Spurverlassenswarnung, Verkehrszeichenassistent und intelligente Scheinwerfersteuerung bereitstellt. Eine oder mehrere der Kamera(s) (z.B. alle Kameras) können simultan Bilddaten (z.B. ein Video) aufnehmen und bereitstellen.

[0074] Eine oder mehrere der Kameras können in einer Montagebaugruppe, z.B. einer kundenspezifisch entworfenen (dreidimensional (3D)-gedruckten) Baugruppe, montiert sein, um Streulicht und Reflexionen aus dem Inneren des Autos (z.B. Reflexionen vom Armaturenbrett, die sich in den Windschutzscheibenspiegeln spiegeln) auszuschließen, welche die Bilddatenerfassungsfähigkeiten der Kamera beeinträchtigen können. Unter Bezugnahme auf Seitenspiegelmontagebaugruppen können die Seitenspiegelbaugruppen kundenspezifisch 3D-gedruckt werden, sodass die Kameramontageplatte mit der Form des Seitenspiegels übereinstimmt. In manchen Beispielen kann die Kamera(s) in den Seitenspiegel integriert sein. Bei Seitensichtkameras können die Kamera(s) auch in vier Säulen an jeder Ecke des Fahrerhauses integriert sein.

[0075] Kameras mit einem Sichtfeld, das Teile der Umgebung vor dem Fahrzeug 900 umfasst (z.B. Frontkameras), können für die Surround-Ansicht verwendet werden, um nach vorne gerichtete Wege und Hindernisse zu identifizieren und dabei zu helfen, mithilfe eines oder mehrerer Controller 936 und/oder Steuer-SoCs, die wichtigen Informationen zur Generierung eines Belegungsrasters und/oder zur Bestimmung der bevorzugten Fahrzeugwege bereitzustellen. Nach vorn gerichtete Kameras können verwendet werden, um viele der gleichen ADAS-Funktionen wie LIDAR auszuführen, einschließlich Notbremsung, Fußgängererkennung und Kollisionsvermeidung. Nach vorn gerichtete Kameras können auch für ADAS-Funktionen und -Systeme verwendet werden, einschließlich Spurverlassenswarnungen (Lane Departure Warning - LDW), autonome Geschwindigkeitssteuerung (Autonomous Cruise Control - ACC) und/oder andere Funktionen wie Verkehrszeichenerkennung.

[0076] Eine Vielfalt an Kameras kann in einer nach vorn gerichteten Konfiguration verwendet werden, einschließlich zum Beispiel einer monokularen Kameraplattform, die einen Farbbildsensor mit CMOS (complementary metal oxide semiconductor - komplementärer Metalloxid-Halbleiter) beinhaltet. Ein weiteres Beispiel sind/ist eine Weitwinkelka-

mera(s) 970, die verwendet werden können/kann, um Objekte wahrzunehmen, die aus der Peripherie in den Blick kommen (z.B. Fußgänger, Kreuzungsverkehr oder Fahrräder). Obwohl nur eine Weitwinkelkamera in **Fig. 9B** dargestellt ist, kann es eine beliebige Anzahl von Weitwinkelkameras 970 auf dem Fahrzeug 900 geben. Zusätzlich kann eine/können Long-Range-Kamera(s) 998 (z.B. ein Langbild-Stereokamerapaar) für die tiefenbasierte Objekterkennung verwendet werden, insbesondere für Objekte, für die noch kein neuronales Netzwerk trainiert wurde. Die Long-Range-Kamera(s) 998 kann/können auch zur Objekterkennung und -klassifizierung sowie zur einfachen Objektverfolgung verwendet werden.

[0077] Eine beliebige Anzahl von Stereokameras 968 kann auch in einer nach vorne gerichteten Konfiguration enthalten sein. In zumindest einer Ausführungsform kann/können eine oder mehrere Stereokamera(s) 968 eine integrierte Steuereinheit mit einer skalierbaren Verarbeitungseinheit enthalten, die eine programmierbare Logik (FPGA) und einen Multi-core-Mikroprozessor mit einer integrierten CAN (Controller Area Network)- oder Ethernet-Schnittstelle auf einem einzigen Chip bereitstellen kann. Eine solche Einheit kann verwendet werden, um eine 3D-Karte der Umgebung des Fahrzeugs zu erzeugen, die eine Abstandsschätzung für alle Punkte im Bild beinhaltet. (Eine) alternative Stereokamera(s) 968 kann/können einen/mehrere kompakte Stereo-Vision-Sensor(en) enthalten, der/die zwei Kameraobjektive (jeweils eines links und rechts) und einen Bildverarbeitungs-Chip umfassen kann/können, der den Abstand vom Fahrzeug zum Zielobjekt misst und die generierten Informationen (z.B. Metadaten) verwendet, um die autonomen Notbremsfunktion und den Spurhalteassistent zu aktivieren. Andere Stereokameratypen 968 können zusätzlich oder alternativ zu den hier beschriebenen verwendet werden.

[0078] Kameras mit einem Sichtfeld, das Teile der Umgebung an der Seite des Fahrzeugs 900 umfasst (z.B. Seitenansicht-Kameras), können für die Surround-Ansicht verwendet werden und Informationen liefern, um das Belegungsraaster zu erzeugen und zu aktualisieren sowie Seitenkollisionswarnungen zu erzeugen. Zum Beispiel kann/können Surround-Kamera(s) 974 (z.B. vier Surround-Kameras 974, wie in **Fig. 9B** dargestellt) auf dem Fahrzeug 900 positioniert werden. Die Surround-Kamera(s) 974 kann/können Weitwinkelkameras 970, Fischaugenkameras, 360-Grad-Kameras und/oder ähnliche Kameras umfassen. Zum Beispiel können vier Fischaugenkameras vorne, hinten und an den Seiten des Fahrzeugs angebracht werden. In einer alternativen Anordnung kann das Fahrzeug drei Surround-Kameras 974 (z.B. links, rechts und hinten) verwenden und eine oder mehrere andere Kameras (z.B. eine nach

vorne gerichtete Kamera) als vierte Surround-View-Kamera nutzen.

[0079] Kameras mit einem Sichtfeld, das Teile der Umgebung im Heck des Fahrzeugs 900 umfasst (z.B. Rückfahrkameras), können zur Parkhilfe, zur Surround-Ansicht, zur Warnung vor Heckkollisionen sowie zur Erstellung und Aktualisierung des Belegungsraasters verwendet werden. Es kann eine Vielzahl von Kameras verwendet werden, einschließlich, aber nicht beschränkt auf, Kameras, die auch als Frontkameras geeignet sind (z.B. Long-Range und/oder Mid-Range-Kameras 998, Stereokameras 968, Infrarotkameras 972 usw.), wie hier beschrieben.

[0080] **Fig. 9C** ist ein Blockdiagramm einer beispielhaften Systemarchitektur für das beispielhafte autonome Fahrzeug 900 aus **Fig. 9A**, gemäß manchen Ausführungsformen der vorliegenden Offenbarung. Es versteht sich, dass diese und andere hierin beschriebene Anordnungen nur als Beispiele aufgeführt werden. Andere Anordnungen und Elemente (z.B. Maschinen, Schnittstellen, Funktionen, Befehle, Gruppierungen von Funktionen usw.) können zusätzlich oder anstelle der gezeigten verwendet werden, und manche Elemente können ganz weggelassen werden. Ferner sind viele der hierin beschriebenen Elemente funktionale Einheiten, die als diskrete oder verteilte Komponenten oder in Verbindung mit anderen Komponenten und in jeder geeigneten Kombination und Position umgesetzt werden können. Verschiedene hierin als von Einheiten ausgeführt beschriebene Funktionen können durch Hardware, Firmware und/oder Software ausgeführt werden. Zum Beispiel können verschiedene Funktionen durch einen Prozessor ausgeführt werden, der im Speicher gespeicherte Anweisungen ausführt.

[0081] Alle Komponenten, Merkmale und Systeme des Fahrzeugs 900 in **Fig. 9C** sind als über Bus 902 verbunden dargestellt. Der Bus 902 kann eine Controller Area Network-Datenschnittstelle (CAN) umfassen (alternativ auch hier als „CAN-Bus“ bezeichnet). Ein CAN kann ein Netzwerk innerhalb des Fahrzeugs 900 sein, das zur Unterstützung der Steuerung verschiedener Merkmale und Funktionalitäten des Fahrzeugs 900 verwendet wird, wie z.B. Betätigung der Bremsen, Beschleunigung, Bremsen, Lenken, Scheibenwischer usw. Ein CAN-Bus kann dazu konfiguriert sein, dass er Dutzende oder sogar Hunderte von Knoten mit jeweils einer eigenen eindeutigen Kennung aufweist (z.B. eine CAN-ID). Der CAN-Bus kann ausgelesen werden, um Lenkradwinkel, Geschwindigkeit über Grund, Motorumdrehungen pro Minute (revolutions per minute - RPMs), Tastenpositionen und/oder andere Fahrzeugstatusindikatoren zu bestimmen. Der CAN-Bus kann ASIL B-konform sein.

[0082] Obwohl der Bus 902 hier als CAN-Bus beschrieben wird, soll dies nicht einschränkend sein. Zum Beispiel können zusätzlich zu oder alternativ zum CAN-Bus auch FlexRay und/oder Ethernet verwendet werden. Auch dies soll, obwohl eine einzelne Leitung verwendet wird, um den Bus 902 darzustellen, nicht einschränkend sein. Zum Beispiel kann es eine beliebige Anzahl von Bussen 902 geben, die einen oder mehrere CAN-Busse, einen oder mehrere FlexRay-Busse, einen oder mehrere Ethernet-Busse und/oder einen oder mehrere andere Bustypen mit einem anderen Protokoll umfassen können. In manchen Beispielen können zwei oder mehr Busse 902 verwendet werden, um verschiedene Funktionen auszuführen und/oder um Redundanz sicherzustellen. Zum Beispiel kann ein erster Bus 902 zur Kollisionsvermeidung und ein zweiter Bus 902 zur Ansteuerung verwendet werden. In jedem Beispiel kann jeder Bus 902 mit einer beliebigen Komponente des Fahrzeugs 900 kommunizieren, und zwei oder mehr Busse 902 können mit denselben Komponenten kommunizieren. In manchen Beispielen kann jeder SoC 904, jeder Controller 936 und/oder jeder Computer im Fahrzeug auf dieselben Eingabedaten zugreifen (z.B. Eingaben von Sensoren des Fahrzeugs 900) und an einen gemeinsamen Bus, wie den CAN-Bus, angeschlossen werden.

[0083] Das Fahrzeug 900 kann einen oder mehrere Controller 936 umfassen, wie hier in Bezug auf **Fig. 9A** beschrieben wird. Der/die Controller 936 kann/können für eine Vielzahl von Funktionen verwendet werden. Der/die Controller 936 kann/können an eine der verschiedenen anderen Komponenten und Systeme des Fahrzeugs 900 gekoppelt werden und kann/können zur Steuerung des Fahrzeugs 900, der künstlichen Intelligenz des Fahrzeugs 900, des Infotainments für das Fahrzeug 900 und/oder dergleichen verwendet werden.

[0084] Das Fahrzeug 900 kann ein Ein-Chip-System (system on a chip - SoC) 904 umfassen. Das SoC 904 kann CPU 906, GPU 908, Prozessor(en) 910, Cache(s) 912, Beschleuniger 914, Datenspeicher 916 und/oder andere Komponenten und Funktionen, die nicht abgebildet sind, umfassen. Das SoC 904 kann zur Steuerung des Fahrzeugs 900 in einer Vielfalt von Plattformen und Systemen verwendet werden. Zum Beispiel kann das SoC 904 in einem System (z.B. dem System des Fahrzeugs 900) mit einer HD-Karte 922 kombiniert werden, die Kartenaktualisierungen und/oder Aktualisierungen über eine Netzwerkschnittstelle 924 von einem oder mehreren Servern (z.B. Server(n) 978 aus **Fig. 9D**) erhalten können.

[0085] Die CPU(s) 906 kann/können einen CPU-Cluster oder einen CPU-Komplex enthalten (hier alternativ auch als „CCPLEX“ bezeichnet). Die CPU(s) 906 kann/können mehrere Kerne und/oder L2-

Caches enthalten. In manchen Ausführungsformen kann/können die CPU(s) 906 zum Beispiel acht Kerne in einer kohärenten Multi-Prozessor-Konfiguration enthalten. In manchen Ausführungsformen kann/können die CPU(s) 906 vier Dual-Core-Cluster umfassen, wobei jeder Cluster über einen dedizierten L2-Cache verfügt (z.B. einen 2 MB L2-Cache). Die CPU(s) 906 (z.B. CCPLEX) kann/können konfiguriert sein, um den gleichzeitigen Clusterbetrieb zu unterstützen, so dass eine beliebige Kombination des/der CPU(s) 906-Cluster zu einem bestimmten Zeitpunkt aktiv sein kann/können.

[0086] Die CPU(s) 906 kann/können Energieverwaltungsfunktionen umsetzen, die eine oder mehrere der folgenden Merkmale umfassen: einzelne Hardwareblöcke können automatisch im Leerlauf taktgesteuert werden, um dynamische Energie zu sparen; jeder Kerntakt kann gesteuert werden, wenn der Kern aufgrund der Ausführung von WFI/WFE-Befehlen keine aktiven Anweisungen ausführt; jeder Kern kann unabhängig leistungsgesteuert werden; jeder Kern-Cluster kann unabhängig taktgesteuert werden, wenn alle Kerne taktgesteuert oder leistungsgesteuert sind; und/oder jeder Kern-Cluster kann unabhängig leistungsgesteuert werden, wenn alle Kerne leistungsgesteuert sind. Die CPU(s) 906 kann/können einen erweiterten Algorithmus zur Verwaltung von Energiezuständen umsetzen, bei dem zulässige Energiezustände und erwartete Aktivierungszeiten angegeben werden und der Hardware-/Mikrocode den besten Energiezustand für den Kern, Cluster und CCPLEX bestimmt. Die Verarbeitungskerne können vereinfachte Leistungsstatus-Eintragssequenzen in der Software unterstützen, wobei die Arbeit in den Mikrocode ausgelagert wird.

[0087] Die GPU(s) 908 kann/können eine integrierte GPU enthalten (hier alternativ auch als „iGPU“ bezeichnet). Die GPU(s) 908 kann/können programmierbar sein und für parallele Workloads effizient sein. Die GPU(s) 908 kann/können in manchen Beispielen einen erweiterten Tensor-Befehlssatz verwenden. Die GPU(s) 908 kann/können einen oder mehrere Streaming-Mikroprozessoren enthalten, wobei jeder Streaming-Mikroprozessor einen L1-Cache umfassen kann (z.B. einen L1-Cache mit zumindest 96 KB Speicherkapazität) und zwei oder mehr der Streaming-Mikroprozessoren sich einen L2-Cache teilen können (z.B. einen L2-Cache mit einer Speicherkapazität von 512 KB). In manchen Ausführungsformen kann/können die GPU(s) 908 zumindest acht Streaming-Mikroprozessoren enthalten. Die GPU(s) 908 kann/können Programmierschnittstellen für Computeranwendungen (application programming interface - API) verwenden. Außerdem kann/können die GPU(s) 908 eine oder mehrere parallele Rechnerplattformen und/oder Programmiermodelle (z.B. NVIDIA von CUDA) verwenden.

[0088] Die GPU(s) 908 kann/können für die beste Leistung im Automobil- und Embedded-Anwendungsfall leistungsoptimiert sein. Zum Beispiel kann/können die GPU(s) 908 auf einem Fin-Feldeffekttransistor (Fin field-effect transistor - FinFET) hergestellt werden. Dies soll jedoch nicht einschränkend sein und die GPU(s) 908 kann/können mithilfe anderer Halbleiterfertigungsprozesse hergestellt werden. Jeder Streaming-Mikroprozessor kann eine Anzahl von Verarbeitungskernen mit gemischter Genauigkeit beinhalten, die in mehrere Blöcke partitioniert sind. Zum Beispiel, und ohne Einschränkung, könnten 64 PF32-Kerne und 32 PF64-Kerne in vier Verarbeitungsblöcke partitioniert sein. In solch einem Beispiel könnten jedem Verarbeitungsblock 16 FP32-Kerne, 8 FP64-Kerne, 16 INT32-Kerne, zwei NVIDIA TENSOR COREs mit gemischter Genauigkeit für Deep-Learning-Matrixarithmetik, ein L0-Anweisungs-Cache, ein Warp-Planer, eine Verteilungseinheit und/oder eine Registerdatei mit 64 KB zugewiesen sein. Zusätzlich können Streaming-Mikroprozessoren unabhängige parallele Integer- und Fließkomma-Datenwege beinhalten, um eine effiziente Ausführung von Arbeitslasten mit einer Mischung aus Berechnung und Adressierungsrechnungen zu ermöglichen. Die Streaming-Mikroprozessoren können eine unabhängige Thread-Planungsfunktion beinhalten, um eine feinkörnigere Synchronisation und Kooperation zwischen parallelen Threads zu ermöglichen. Die Streaming-Mikroprozessoren können eine Einheit aus kombiniertem L1-Daten-Cache und gemeinsam genutztem Speicher beinhalten, um die Performance zu verbessern, während die Programmierung vereinfacht wird.

[0089] Die GPU(s) 908 kann/können einen Speicher mit hoher Bandbreite (High Bandwidth Memory - HBM) und/oder ein 16 GB HBM2-Speichersubsystem enthalten, um in manchen Beispielen eine Spitzenspeicherbandbreite von etwa 900 GB/s zu bieten. In manchen Beispielen kann zusätzlich oder alternativ zu dem HBM-Speicher ein synchroner Grafik-Direktzugriffsspeicher (synchronous graphics random-access memory - SGRAM) verwendet werden, wie etwa ein synchroner Direktzugriffsspeicher vom Graphics-Double-Data-Rate-Typ fünf (graphics double data rate type five - GDDR5).

[0090] Die GPU(s) 908 kann/können eine Unified-Memory-Technologie einschließlich Zugriffsprozessoren enthalten, um eine genauere Migration von Speicherseiten auf den Prozessor zu ermöglichen, der am häufigsten auf sie zugreift, wodurch die Effizienz für Speicherbereiche, die zwischen Prozessoren gemeinsam genutzt werden, verbessert wird. In manchen Beispielen kann die Unterstützung von Address Translation Services (ATS) verwendet werden, um der/den GPU(s) 908 den direkten Zugriff auf die CPU(s) 906-Seitentabellen zu ermöglichen. In solchen Beispielen kann eine Anfrage zur Adress-

übersetzung an die CPU(s) 906 gesendet werden, wenn die Speicherverwaltungseinheit (memory management unit - MMU) der GPU(s) 908 einen Fehler aufweist. Als Reaktion darauf sucht/suchen die CPU(s) 906 möglicherweise in ihren Seitentabellen nach dem Virtuellen-zu-Physisch-Mapping für die Adresse und überträgt/übertragen die Übersetzung zurück an die GPU(s) 908. Daher kann die Unified-Memory-Technologie einen einzigen einheitlichen virtuellen Adressraum für den Speicher sowohl der CPU(s) 906 als auch der GPU(s) 908 ermöglichen, wodurch die GPU(s) 908-Programmierung und die Portierung von Anwendungen auf die GPU(s) 908 vereinfacht wird.

[0091] Darüber hinaus kann/können die GPU(s) 908 einen Zugriffszähler enthalten, der die Zugriffshäufigkeit der GPU(s) 908 auf den Speicher anderer Prozessoren nachverfolgen kann. Der Zugriffszähler kann dazu beitragen, dass Speicherseiten in den physischen Speicher desjenigen Prozessors verschoben werden, der am häufigsten auf die Seiten zugreift.

[0092] Das SoC 904 kann eine beliebige Anzahl von Cache(s) 912 enthalten, einschließlich der hier beschriebenen. Der Cache(s) 912 kann zum Beispiel einen L3-Cache enthalten, der sowohl der/den CPU(s) 906 als auch der/den GPU(s) 908 zur Verfügung steht (z.B. der sowohl an die CPU(s) 906 als auch die GPU(s) 908 angeschlossen ist). Der/die Cache(s) 912 kann/können einen Write-Back-Cache enthalten, der den Status von Zeilen nachverfolgen kann, z.B. durch Verwendung eines Cache-Kohärenzprotokolls (z.B. MEI, MESI, MSI usw.). Der L3-Cache kann in Abhängigkeit von der Ausführungsform 4 MB oder mehr beinhalten, obwohl auch kleinere Cache-Größen verwendet werden können.

[0093] Der/die SoC(s) 904 kann/können eine arithmetische Logikeinheit(en) (ALU(s)) beinhalten, die bei der Durchführung von Verarbeitungen in Bezug auf eine der vielen Aufgaben oder Operationen des Fahrzeugs 900 - wie z.B. der Verarbeitung von DNNs - genutzt werden kann. Darüber hinaus können die SoC(s) 904 eine oder mehrere Gleitkommaeinheiten (floating point units - FPU(s)) - oder andere mathematische Coprozessoren oder numerische Coprozessoren - zur Durchführung mathematischer Operationen innerhalb des Systems beinhalten. Zum Beispiel können die SoC(s) 104 eine oder mehrere FPUs beinhalten, die als Ausführungseinheiten in einer oder mehreren CPU(s) 906 und/oder GPU(s) 908 integriert sind.

[0094] Das SoC 904 kann einen oder mehrere Beschleuniger 914 enthalten (z.B. Hardware-Beschleuniger, Software-Beschleuniger oder eine Kombination davon). Zum Beispiel kann das SoC 904 einen Hardware-Beschleunigungscluster enthal-

ten, der optimierte Hardware-Beschleuniger und/oder einen großen On-Chip-Speicher umfassen kann. Der große chipinterne Speicher (z.B. 4 MB SRAM) kann einen Hardware-Beschleunigungscluster zur Beschleunigung neuronaler Netzwerke und anderer Berechnungen ermöglichen. Der Hardware-Beschleunigungscluster kann verwendet werden, um die GPU(s) 908 zu ergänzen und manche der Aufgaben der GPU(s) 908 zu entlasten (z.B. um mehr Zyklen der GPU(s) 908 für die Durchführung anderer Aufgaben freizumachen). Als Beispiel können die Beschleuniger 914 für gezielte Workloads (z.B. Wahrnehmung, konvolutionale neuronale Netzwerke (CNN) usw.) verwendet werden, die stabil genug sind, um für eine Beschleunigung geeignet zu sein. Der Begriff „CNN“, wie er hier verwendet wird, kann alle Arten von CNNs beinhalten, einschließlich regionenbasierter oder regionaler neuronaler Faltungnetzwerke (RCNNs) und Fast RCNNs (z.B. für die Objekterkennung).

[0095] Der/die Beschleuniger 914 (z.B. der Hardware-Beschleunigungscluster) kann/können einen/mehrere Deep-Learning-Beschleuniger (deep learning accelerator - DLA) enthalten. Die DLA(s) können eine oder mehrere Tensor-Verarbeitungseinheiten (TPU(s)) beinhalten, die derart konfiguriert sein können, dass sie zusätzliche zehn Billionen Operationen pro Sekunde für Deep-Learning-Anwendungen und -Inferenzierung bereitstellen. Die TPUs können Beschleuniger sein, die zum Durchführen von Bildverarbeitungsfunktionen (z.B. für CNNs, RCNNs usw.) konfiguriert und optimiert sind. Die DLA(s) können ferner für einen spezifischen Satz von Arten von neuronalen Netzwerken und Fließkommaoperationen sowie für die Inferenzierung optimiert sein. Das Design der DLA(s) kann mehr Performance pro Millimeter bereitstellen als eine typische Universal-GPU und übertrifft die Performance einer CPU bei weitem. Die TPU(s) können mehrere Funktionen durchführen, einschließlich einer Einzelinstanz-Faltungsfunktion, die z.B. INT8-, INT16- und FP16-Datenarten sowohl für Merkmale als auch für Gewichtungen unterstützt, sowie Postprozessorfunktionen.

[0096] Der/die DLA(s) können schnell und effizient neuronale Netzwerke, insbesondere CNNs, auf bearbeiteten oder unbearbeiteten Daten für eine Vielzahl von Funktionen ausführen, einschließlich zum Beispiel und ohne Einschränkung: Ein CNN zur Objektidentifizierung und -erkennung mithilfe von Daten von Kamerasensoren; ein CNN für die Abstandsabschätzung mit Daten von Kamerasensoren, ein CNN für die Erkennung und Identifizierung von Rettungsfahrzeugen und Erkennung mit Daten von Mikrofonen; ein CNN für die Gesichtserkennung und die Identifizierung von Fahrzeugbesitzern durch Verwendung von Daten von Kamerasensoren und/o-

der ein CNN für Sicherheits- und/oder sicherheitsrelevante Ereignisse.

[0097] Der/die DLA(s) kann/können jede Funktion der GPU(s) 908 ausführen, und durch die Verwendung eines Inferenzbeschleunigers kann ein Entwickler zum Beispiel für jede Funktion entweder den/die DLA(s) oder die GPU(s) 908 als Ziel verwenden. So kann der Designer zum Beispiel die Verarbeitung von CNNs und Gleitkommaoperationen auf den/die DLA(s) konzentrieren und andere Funktionen der/den GPU(s) 908 und/oder anderen Beschleunigern 914 überlassen.

[0098] Der/die Beschleuniger 914 (z.B. der Hardware-Beschleunigungscluster) können einen/mehrere programmierbare(n) Bildverarbeitungsbeschleuniger (programmable vision accelerator - PVA) enthalten, der/die in diesem Dokument hilfsweise als Bildverarbeitungsbeschleuniger bezeichnet werden kann/können. Die PVA(s) können zur Beschleunigung von Algorithmen des maschinellen Sehens für weiterentwickelte Fahrerassistenzsysteme (ADAS), autonomes Fahren, Augmented-Reality(AR)- und/oder Virtual-Reality(VR)-Anwendungen ausgelegt und konfiguriert sein. Die PVA(s) können ein Gleichgewicht zwischen Performance und Flexibilität bereitstellen. Beispielsweise und ohne Einschränkung können alle PVA(s) eine beliebige Anzahl von Reduced-Instruction-Set-Computer (RISC)-Kerne, direkten Speicherzugriff (direct memory access - DMA) und/oder eine beliebige Anzahl von Vektorprozessoren beinhalten.

[0099] Die RISC-Kerne können mit Bildsensoren (z.B. den Bildsensoren einer beliebigen der hierin beschriebenen Kameras), Bildsignalprozessor(en) und/oder dergleichen interagieren. Jeder der RISC-Kerne kann eine beliebige Menge an Speicher beinhalten. Die RISC-Kerne können in Abhängigkeit von der Ausführungsform ein beliebiges von einer Anzahl von Protokollen verwenden. In manchen Beispielen können die RISC-Kerne ein Echtzeitbetriebssystem (real-time operating system - RTOS) ausführen. Die RISC-Kerne können unter Verwendung einer oder mehrerer Vorrichtungen für integrierte Schaltungen, anwendungsspezifischer integrierter Schaltungen (ASICs) und/oder Speichervorrichtungen umgesetzt sein. Die RISC-Kerne können zum Beispiel einen Anweisungs-Cache und/oder einen eng gekoppelten RAM beinhalten.

[0100] Der DMA kann Komponenten des/der PVA(s) unabhängig von der/den CPU(s) 906 den Zugriff auf den Systemspeicher ermöglichen. Der DMA kann eine beliebige Anzahl von Merkmalen unterstützen, die zum Bereitstellen der Optimierung des PVA verwendet werden, einschließlich der Unterstützung von mehrdimensionaler Adressierung und/oder zirkulärer Adressierung, ohne darauf beschränkt zu sein. In

manchen Beispielen kann der DMA bis zu sechs oder mehr Dimensionen der Adressierung unterstützen, die Blockbreite, Blockhöhe, Blocktiefe, horizontale Blockabstufung, vertikale Blockabstufung und/oder Tiefenabstufung beinhalten können.

[0101] Die Vektorprozessoren können programmierbare Prozessoren sein, die so ausgestaltet sein können, dass sie die Programmierung für Algorithmen des maschinellen Sehens effizient und flexibel ausführen und Signalverarbeitungsfähigkeiten bereitstellen. In manchen Beispielen kann der PVA einen PVA-Kern und zwei Vektorverarbeitungsteil-systempartitionen beinhalten. Der PVA-Kern kann ein Prozessorteilsystem, DMA-Engine(s) (z.B. zwei DMA-Engines) und/oder andere Peripheriegeräte beinhalten. Das Vektorverarbeitungsteilsystem kann als primäre Verarbeitungs-Engine des PVA arbeiten und kann eine Vektorverarbeitungseinheit (vector processing unit - VPU), einen Anweisungs-Cache und/oder einen Vektorspeicher (z.B. VMEM) beinhalten. Ein VPU-Kern kann einen digitalen Signalprozessor beinhalten, wie zum Beispiel einen digitalen Single-Instruction-Multiple-Data-(SIMD-)Very-Long-Instruction-Word-(VLIW-)Signalprozessor. Die Kombination von SIMD und VLIW kann den Durchsatz und die Geschwindigkeit erhöhen.

[0102] Jeder der Vektorprozessoren kann einen Anweisungs-Cache beinhalten und an dedizierten Speicher gekoppelt sein. Daher kann in manchen Beispielen jeder der Vektorprozessoren derart konfiguriert sein, dass er unabhängig von den anderen Vektorprozessoren ausgeführt wird. In anderen Beispielen können Vektorprozessoren, die in einem konkreten PVA enthalten sind, derart konfiguriert sein, dass sie Datenparallelität einsetzen. Zum Beispiel kann in manchen Ausführungsformen die Vielzahl von Vektorprozessoren, die in einem einzelnen PVA enthalten ist, denselben Algorithmus des maschinellen Sehens ausführen, jedoch an unterschiedlichen Regionen eines Bildes. In anderen Beispielen können die in einem konkreten PVA enthaltenen Vektorprozessoren simultan unterschiedliche Algorithmen des maschinellen Sehens an demselben Bild ausführen oder sogar unterschiedliche Algorithmen an sequentiellen Bildern oder Abschnitten eines Bildes ausführen. Unter anderem kann eine beliebige Anzahl von PVAs in dem Hardware-Beschleunigungscluster enthalten sein und kann eine beliebige Anzahl von Vektorprozessoren in jedem der PVAs enthalten sein. Zusätzlich können der/die PVA(s) einen zusätzlichen Fehlerkorrekturcode(ECC)-Speicher beinhalten, um die Gesamtsystemsicherheit zu erhöhen.

[0103] Der/die Beschleuniger 914 (z.B. der Hardware-Beschleunigungscluster) können ein Computervision-Netzwerk auf dem Chip und SRAM enthalten, um einen SRAM mit hoher Bandbreite und

niedriger Latenz für den/die Beschleuniger 914 bereitzustellen. In manchen Beispielen kann der chipinterne Speicher zumindest 4 MB SRAM beinhalten, der z.B. und ohne Einschränkung aus acht feldkonfigurierbaren Speicherblöcken besteht, auf die sowohl der PVA als auch der DLA zugreifen können. Jedes Paar von Speicherblöcken kann eine weiterentwickelte Peripheriebus (advanced peripheral bus - APB)-Schnittstelle, eine Konfigurationsschaltung, eine Steuerung und einen Multiplexer beinhalten. Es kann jede Art von Speicher verwendet werden. Der PVA und DLA können auf den Speicher über einen Backbone zugreifen, der dem PVA und DLA einen Hochgeschwindigkeitszugriff auf den Speicher bereitstellt. Der Backbone kann ein Netzwerk auf dem Chip für maschinelles Sehen beinhalten, das den PVA und DLA mit dem Speicher verbindet (z.B. unter Verwendung von dem APB).

[0104] Das Netzwerk kann auf dem Chip für maschinelles Sehen eine Schnittstelle beinhalten, die vor der Übertragung eines beliebigen Steuersignals/einer beliebigen Adresse/ beliebiger Daten bestimmt, dass sowohl der PVA als auch der DLA einsatzbereite und gültige Signale bereitstellen. Eine derartige Schnittstelle kann separate Phasen und separate Kanäle für die Übertragung von Steuersignalen/Adressen/Daten sowie eine Burst-artige Kommunikation für eine kontinuierliche Datenübertragung bereitstellen. Diese Art von Schnittstelle kann den Normen ISO 26262 oder IEC 61508 entsprechen, obwohl auch andere Normen und Protokolle verwendet werden können.

[0105] In manchen Beispielen kann das SoC 904 einen Echtzeitstrahlverfolgungs-Hardware-Beschleuniger beinhalten, wie in der am 10. August 2018 eingereichten US-Patentanmeldung Nr. 16/101,232 beschrieben. Der Echtzeitstrahlverfolgungs-Hardware-Beschleuniger kann verwendet werden, um schnell und effizient Positionen und Ausdehnungen von Objekten (z.B. innerhalb eines Weltmodells) zu bestimmen, um Echtzeitvisualisierungssimulationen zu erzeugen, für die RADAR-Signalinterpretation, für die Schallausbreitungssynthese und/oder -analyse, für die Simulation von SONAR-Systemen, für die allgemeine Wellenausbreitungssimulation, für den Vergleich mit LIDAR-Daten zum Zwecke der Lokalisierung und/oder für andere Funktionen und/oder für andere Anwendungen. In manchen Ausführungsformen können eine oder mehrere Tree Traversal Units (TTUs) für die Ausführung einer oder mehrerer Operationen verwendet werden, die mit der Strahlverfolgung in Verbindung stehen.

[0106] Der/die Beschleuniger 914 (z.B. der Hardware-Beschleunigercluster) haben ein breites Spektrum von Einsatzmöglichkeiten für autonomes Fahren. Der PVA kann ein programmierbarer

Sichtbeschleuniger sein, der für wichtige Verarbeitungsstufen im ADAS und in autonomen Fahrzeugen verwendet werden kann. Die Fähigkeiten des PVA sind eine gute Ergänzung für algorithmische Domänen, die eine vorhersagbare Verarbeitung bei niedriger Leistung und niedriger Latenz benötigen. Anders ausgedrückt zeigt der PVA eine gute Performance für halbdichte oder dichte reguläre Berechnungen, auch an kleinen Datensätzen, die vorhersagbare Laufzeiten mit niedriger Latenz und niedriger Leistung benötigen. Folglich sind die PVAs im Zusammenhang mit Plattformen für autonome Fahrzeuge für die Ausführung klassischer Algorithmen für maschinelles Sehen konstruiert, da diese effizient bei der Objekterkennung sind und mit Integer-Mathematik arbeiten.

[0107] Zum Beispiel wird gemäß einer Ausführungsform der Technologie der PVA verwendet, um maschinelles Stereo-Sehen durchzuführen. Ein auf semiglobalem Abgleich basierender Algorithmus kann verwendet werden, obwohl dies nicht als Einschränkung auszulegen ist. Viele Anwendungen für das autonome Fahren auf Level 3-5 erfordern Bewegungsschätzung/Stereo-Abgleich spontan (z.B. Struktur aus Bewegung, Fußgängererkennung, Fahrspurerkennung usw.). Der PVA kann eine Funktion des maschinellen Stereo-Sehens an Eingaben von zwei monokularen Kameras durchführen.

[0108] In manchen Beispielen kann der PVA verwendet werden, um einen dichten optischen Fluss durchzuführen. Gemäß der Verarbeitung von RADAR-Rohdaten (z.B. mit einer 4D-Fast-Fourier-Transformation) zur Bereitstellung von verarbeitetem RADAR. In anderen Beispielen wird der PVA für die Laufzeit-Tiefenverarbeitung verwendet, indem z.B. Laufzeit-Rohdaten verarbeitet werden, um verarbeitete Laufzeitdaten bereitzustellen.

[0109] Der DLA kann verwendet werden, um eine beliebige Art von Netzwerk auszuführen, um die Steuerung und Fahrsicherheit zu verbessern, einschließlich zum Beispiel ein neuronales Netz, das ein Maß an Konfidenz für jede Objekterkennung ausgibt. Ein derartiger Konfidenzwert kann als eine Wahrscheinlichkeit interpretiert werden oder als Bereitstellung einer relativen „Gewichtung“ jeder Erkennung im Vergleich zu anderen Erkennungen. Der Konfidenzwert ermöglicht es dem System, weitere Bestimmungen darüber zu treffen, welche Erkennungen als richtig positive Erkennungen und nicht als falsch positive Erkennungen betrachtet werden sollten. Zum Beispiel kann das System einen Schwellenwert für die Konfidenz festlegen und nur Erkennungen, die den Schwellenwert überschreiten, als richtig positive Erkennungen betrachten. In einem automatischen Notbrems-(automatic emergency braking - AEB-)System würden falsch positive Erkennungen dazu führen, dass das Fahrzeug automatisch eine Notbremsung durchführt, was natürlich

unerwünscht ist. Daher sollten nur die sichersten Erkennungen als Auslöser für AEB in Betracht gezogen werden. Der DLA kann ein neuronales Netzwerk zur Regression des Konfidenzwerts ausführen. Das neuronale Netzwerk kann als Eingabe zumindest eine Teilmenge von Parametern verwenden, wie z.B. Abmessungen des Begrenzungsrahmens, Schätzung der Bodenebene (z.B. von einem anderen Subsystem), IMU-Sensor-966-Ausgabe (Inertial Measurement Unit - IMU), die mit der Ausrichtung des Fahrzeugs 900, Abstand, 3D-Positionsabschätzungen des Objekts, das aus dem neuronalen Netzwerk und/oder anderen Sensoren (z.B. LIDAR-Sensor(en) 964 oder RADAR-Sensor(en) 960), unter anderem gewonnen wurden, korreliert.

[0110] Das SoC 904 kann Datenspeicher 916 (z.B. Speicher) enthalten. Der Datenspeicher 916 kann ein On-Chip-Speicher des SoC 904 sein, das neuronale Netzwerke speichern kann, die auf der GPU und/oder der DLA ausgeführt werden. In manchen Beispielen kann der Datenspeicher 916 groß genug sein, um mehrere Instanzen neuronaler Netzwerke für Redundanz und Sicherheit zu speichern. Der Datenspeicher 912 kann L2- oder L3-Cache(s) 912 umfassen. Der Verweis auf den Datenspeicher 916 kann, wie hier beschrieben, den Bezug zum Speicher des PVA-, DLA- und/oder anderen Beschleunigern 914 enthalten.

[0111] Das SoC 904 kann einen oder mehr Prozessoren 910 (z.B. eingebettete Prozessoren) enthalten. Die Prozessoren 910 können einen Boot- und Energieverwaltungsprozessor enthalten, der ein dedizierter Prozessor und ein Subsystem sein kann, um die Boot-Energie- und Verwaltungsfunktionen sowie die damit verbundene Sicherheitsdurchsetzung zu handhaben. Der Boot- und Energieverwaltungsprozessor kann Teil der SoC 904-Startsequenz sein und kann Laufzeitenergieverwaltungsdienste bereitstellen. Der Bootenergie- und Verwaltungsprozessor kann die Taktfrequenz- und Spannungsprogrammierung, Unterstützung bei Übergängen des Energiesparzustands des Systems, die Verwaltung von SoC 904-Thermik und Temperatursensoren und/oder die Verwaltung der SoC 904-Leistungszustände bereitstellen. Jeder Temperatursensor kann als Ringoszillator umgesetzt werden, dessen Ausgangsfrequenz proportional zur Temperatur ist, und das SoC 904 kann die Ringoszillatoren verwenden, um Temperaturen von CPU(s) 906, GPU(s) 908 und/oder Beschleunigern 914 zu erkennen. Wenn festgestellt wird, dass die Temperaturen einen Schwellenwert überschreiten, kann der Boot- und Energieverwaltungsprozessor in eine Temperaturfehleroutine wechseln und das SoC 904 in einen niedrigeren Leistungszustand versetzen und/oder das Fahrzeug 900 in einen Sicherer-Stopp-Modus versetzen (z.B. das Fahrzeug 900 zu einem sicheren Halt führen).

[0112] Die Prozessoren 910 können außerdem einen Satz integrierter Prozessoren enthalten, die als Audioprozessor dienen können. Die Audioverarbeitungs-Engine kann ein Audioteilsystem sein, das eine vollständige Hardware-Unterstützung für Mehrkanal-Audio über mehrere Schnittstellen sowie eine breite und flexible Palette von Audio-E/A-Schnittstellen ermöglicht. In manchen Beispielen ist die Audioverarbeitungs-Engine ein dedizierter Prozessorkern mit einem digitalen Signalprozessor mit dediziertem RAM

[0113] Der/die Prozessor(en) 910 kann/können außerdem mit einem „Always-on“-Prozessor ausgestattet sein, der die erforderlichen Hardwarefunktionen zur Unterstützung der Sensorverwaltung mit niedrigem Energieverbrauch und der Anwendungsfälle für das Einschalten bereitstellen kann. Die stets eingeschaltete Prozessor-Engine kann einen Prozessorkern, einen eng gekoppelten RAM, unterstützende Peripheriegeräte (z.B. Timer und Unterbrechungssteuerungen), verschiedene E/A-Steuerungssperipheriegeräte und Routing-Logik beinhalten.

[0114] Der/die Prozessor(en) 910 kann/können außerdem eine Sicherheits-Cluster-Engine enthalten, die ein dediziertes Prozessor-Subsystem für das Sicherheitsmanagement für Automobilanwendungen enthält. Die Sicherheits-Cluster-Engine kann zwei oder mehr Prozessorkerne, einen eng gekoppelten RAM, unterstützende Peripheriegeräte (z.B. Zeitgeber, eine Unterbrechungssteuerung usw.) und/oder Routing-Logik beinhalten. In einem Sicherheitsmodus können die zwei oder mehr Kerne in einem Gleichschrittmodus arbeiten und als ein einzelner Kern mit einer Vergleichslogik funktionieren, um beliebige Unterschiede zwischen ihren Operationen zu erkennen.

[0115] Der/die Prozessor(en) 910 kann/können außerdem einen Echtzeit-Kamera-Engine enthalten, der ein dediziertes Prozessor-Subsystem für die Echtzeitkameraverwaltung enthalten kann.

[0116] Der/die Prozessor(en) 910 kann/können außerdem einen Signalprozessor mit hohem Dynamikbereich enthalten, der einen Bildsignalprozessor enthält, der als Hardware-Engine Teil der Kameraverarbeitungspipeline ist.

[0117] Der/die Prozessor(en) 910 kann/können einen Videobild-Compositor enthalten, der ein Verarbeitungsblock sein kann (z.B. auf einem Mikroprozessor umgesetzt), der Video-Nachbearbeitungsfunktionen umgesetzt, die von einer Videowiedergabeanwendung benötigt werden, um das endgültige Bild für das Playerfenster zu erzeugen. Der Videobild-Compositor kann die Objektivverzerrungskorrektur an Weitwinkelkameras 970, Surround-Kameras 974 und/oder an Sensoren der

Fahrgastraum-Überwachungskamera durchführen. Der kabineninterne Überwachungskamerasensor wird vorzugsweise von einem neuronalen Netzwerk überwacht, das auf einer anderen Instanz des Advanced SoC läuft und derart konfiguriert ist, dass es Ereignisse in der Kabine erkennt und entsprechend reagiert. Ein kabineninternes System kann Lippenlesen durchführen, um den Mobilfunkdienst zu aktivieren und einen Anruf zu tätigen, E-Mails zu diktieren, ein Ziel des Fahrzeugs zu ändern, ein Infotainmentsystem des Fahrzeugs und dessen Einstellungen zu aktivieren oder zu ändern oder sprachaktiviertes Surfen im Internet bereitzustellen. Dem Fahrer stehen bestimmte Funktionen nur zur Verfügung, wenn das Fahrzeug in einem autonomen Modus betrieben wird, und sind ansonsten deaktiviert.

[0118] Der Videobildkompositor kann eine erweiterte zeitliche Rauschunterdrückung sowohl für die räumliche als auch für die zeitliche Rauschunterdrückung beinhalten. Wenn, zum Beispiel Bewegung in einem Video vorkommt, gewichtet die Rauschunterdrückung die räumlichen Informationen entsprechend, indem sie die Gewichtung der Informationen, die von benachbarten Frames bereitgestellt werden, verringert. Wenn ein Bild oder ein Abschnitt eines Bildes keine Bewegung beinhaltet, kann die durch den Videobildkompositor durchgeführte zeitliche Rauschunterdrückung Informationen aus dem vorherigen Bild verwenden, um das Rauschen in einem derzeitigen Bild zu unterdrücken.

[0119] Der Videobild-Compositor kann auch derart konfiguriert werden, dass Stereo-Rektifikation auf Stereoeingangs-Linsenframes durchgeführt wird. Der Videobild-Compositor kann auch für die Erstellung der Benutzeroberfläche verwendet werden, wenn der Betriebssystemdesktop verwendet wird und die GPU(s) 908 nicht erforderlich ist, um kontinuierlich neue Oberflächen zu rendern. Selbst wenn die GPU(s) 908 eingeschaltet ist und aktiv 3D-Rendern macht, kann der Videobild-Compositor verwendet werden, um die GPU(s) 908 zu entlasten, um die Leistung und Reaktionsfähigkeit zu verbessern.

[0120] Das SoC 904 kann außerdem eine serielle Schnittstelle für mobile Industrieprozessorschnittstellen (mobile industry processor interface - MIPI) für den Empfang von Video und Input von Kameras, eine Hochgeschwindigkeitsschnittstelle und/oder einen Videoeingangsblock enthalten, der für Kamera- und verwandte Pixeleingabefunktionen verwendet werden kann. Das SoC 904 kann außerdem eine Eingabe-/Ausgabesteuerung umfassen, die von der Software gesteuert werden kann und für den Empfang von I/O-Signalen verwendet werden kann, die nicht einer bestimmten Rolle zugewiesen sind.

[0121] Das SoC 904 kann darüber hinaus eine breite Palette von Peripherieschnittstellen enthalten, um die Kommunikation mit Peripheriegeräten, Audio-Codecs, Energieverwaltung und/oder anderen Geräten zu ermöglichen. Das SoC 904 kann zur Verarbeitung von Daten von Kameras (z.B. verbunden über Gigabit Multimedia Serial Link und Ethernet), Sensoren (z.B. LIDAR-Sensor(en) 964, RADAR-Sensor(en) 960 usw., die über Ethernet angeschlossen werden können), Daten von Bus 902 (z.B. Geschwindigkeit von Fahrzeug 900, Lenkradposition usw.), Daten von GNSS-Sensor(en) 958 (z.B. über Ethernet oder CAN-Bus angeschlossen), verwendet werden. Das SoC 904 kann außerdem dedizierte Hochleistungs-Massenspeichercontroller enthalten, die ihre eigenen DMA-Engines enthalten können und die dazu verwendet werden können, die CPU(s) 906 von routinemäßigen Datenverwaltungsaufgaben zu befreien.

[0122] Das SoC 904 kann eine End-to-End-Plattform mit einer flexiblen Architektur sein, die die Automatisierungsstufen 3 bis 5 umfasst und somit eine umfassende funktionale Sicherheitsarchitektur bietet, die Computervision und ADAS-Techniken für Vielfalt und Redundanz effizient nutzt, eine Plattform für einen flexiblen, zuverlässigen Fahrsoftware-Stack bietet, zusammen mit Deep-Learning-Tools. Das SoC 904 kann schneller, zuverlässiger und noch energieeffizienter und platzsparender als herkömmliche Systeme sein. So kann der Beschleuniger 914 in Kombination mit der CPU 906, der GPU 908 und dem Datenspeicher 916 eine schnelle, effiziente Plattform für autonome Fahrzeuge der Stufe 3-5 bereitstellen.

[0123] Die Technologie bietet somit Fähigkeiten und Funktionalität, die mit herkömmlichen Systemen nicht erreicht werden können. Zum Beispiel können Algorithmen des maschinellen Sehens auf CPUs ausgeführt werden, die unter Verwendung einer Programmiersprache auf hohem Level, wie z.B. der Programmiersprache C, konfiguriert werden können, um eine große Vielfalt von Verarbeitungsalgorithmen über eine große Vielfalt von visuellen Daten auszuführen. Die CPUs sind jedoch oft nicht in der Lage, die Performance-Anforderungen vieler Anwendungen des maschinellen Sehens zu erfüllen, wie z.B. in Bezug auf die Ausführungszeit und den Leistungsverbrauch. Insbesondere sind viele CPUs nicht in der Lage, komplexe Objekterkennungsalgorithmen in Echtzeit auszuführen, die in fahrzeuginternen ADAS-Anwendungen und in praktischen autonomen Fahrzeugen der Levels 3-5 erforderlich sind.

[0124] Im Gegensatz zu herkömmlichen Systemen ermöglicht die hier beschriebene Technologie durch Bereitstellung eines CPU-Komplexes, eines GPU-Komplexes und eines Hardware-Beschleunigungsklusters die gleichzeitige und/oder sequenzielle

Durchführung mehrerer neuronaler Netzwerke und die Kombination der Ergebnisse, um die Level 3-5-Funktionalität für autonomes Fahren zu ermöglichen. Ein CNN zum Beispiel, das auf der DLA oder dGPU ausgeführt wird (z.B. die GPU 920), kann eine Text- und Worterkennung enthalten, die es dem Supercomputer ermöglicht, Verkehrszeichen zu lesen und zu verstehen, einschließlich Zeichen, für die das neuronale Netzwerk nicht speziell trainiert wurde. Der DLA kann ferner ein neuronales Netzwerk beinhalten, das in der Lage ist, Zeichen zu identifizieren, zu interpretieren und ein semantisches Verständnis davon bereitzustellen und dieses semantische Verständnis an den Wegplanungsmodul weiterzugeben, die auf dem CPU-Komplex laufen.

[0125] Als weiteres Beispiel können mehrere neuronale Netzwerke simultan ausgeführt werden, wie für das Fahren bei Level 3, 4 oder 5 erforderlich ist. Zum Beispiel kann ein Warnschild mit der Aufschrift „Vorsicht: Blinkende Lichter weisen auf Vereisung hin“ zusammen mit einem elektrischen Licht von mehreren neuronalen Netzwerken unabhängig oder gemeinsam interpretiert werden. Das Schild selbst kann von einem ersten eingesetzten neuronalen Netzwerk (z.B. einem trainierten neuronalen Netzwerk) als Verkehrsschild identifiziert werden und kann der Text „Blinkende Lichter weisen auf Vereisung hin“ von einem zweiten eingesetzten neuronalen Netzwerk interpretiert werden, das die Wegplanungssoftware des Fahrzeugs (die vorzugsweise auf dem CPU-Komplex ausgeführt wird) darüber informiert, dass, wenn blinkende Lichter erkannt werden, Vereisungen vorliegen. Das blinkende Licht kann identifiziert werden, indem ein drittes eingesetztes neuronales Netzwerk über mehrere Frames hinweg betrieben wird und die Wegplanungssoftware des Fahrzeugs über das Vorhandensein (oder Nichtvorhandensein) von blinkenden Lichtern informiert. Alle drei neuronalen Netzwerke können gleichzeitig ausgeführt werden, z.B. innerhalb des DLA und/oder auf der GPU 908.

[0126] In manchen Beispielen kann ein CNN zur Gesichtserkennung und Fahrzeugidentifikation Daten von Kamerasensoren verwenden, um die Anwesenheit eines autorisierten Fahrers und/oder Fahrzeugbesitzers zu identifizieren 900. Die stets eingeschaltete Sensorverarbeitungs-Engine kann verwendet werden, um das Fahrzeug zu entriegeln, wenn sich der Besitzer der Fahrertür nähert und Lichter einschaltet, und um in dem Sicherheitsmodus das Fahrzeug zu deaktivieren, wenn der Besitzer das Fahrzeug verlässt. Auf diese Weise sorgen die SoCs 904 für Sicherheit gegen Entwendung und/oder Autodiebstahl.

[0127] In einem anderen Beispiel kann ein CNN zur Erkennung und Identifizierung von Rettungsfahrzeugen Daten von Mikrofonen 996996 verwenden, um

Sirenen von Notfahrzeugen zu erkennen und zu identifizieren. Im Gegensatz zu herkömmlichen Systemen, die allgemeine Klassifikatoren verwenden, um Sirenen zu erkennen und Funktionen manuell zu extrahieren, verwenden die SoCs 904 das CNN zur Klassifizierung von Umwelt- und Stadtgeräuschen sowie zur Klassifizierung von visuellen Daten. In einer bevorzugten Ausführungsform wird das CNN, das auf dem DLA läuft, dafür trainiert, die relative Annäherungsgeschwindigkeit des Einsatzfahrzeugs zu identifizieren (z.B. durch Verwendung des Dopplereffekts). Das CNN kann auch dazu trainiert werden, Einsatzfahrzeuge zu identifizieren, die für den lokalen Einsatzbereich des Fahrzeugs spezifisch sind, wie durch GNSS-Sensor(en) 958 identifiziert. Folglich versucht das CNN zum Beispiel, wenn es in Europa betrieben wird, europäische Sirenen zu erkennen, und in den Vereinigten Staaten versucht das CNN, nur nordamerikanische Sirenen zu identifizieren. Sobald ein Rettungsfahrzeug erkannt wird, kann ein Steuerprogramm verwendet werden, um eine Sicherheitsroutine für Rettungsfahrzeuge auszuführen, das Fahrzeug zu verlangsamen, zur Seite der Straße zu fahren, das Fahrzeug abzustellen und/oder das Fahrzeug mit Hilfe der Ultraschallsensoren 962 im Leerlauf laufen zu lassen, bis das Rettungsfahrzeug vorbeifährt.

[0128] Das Fahrzeug kann eine CPU 918 (z.B. diskrete CPU oder dCPU) enthalten, die über eine Hochgeschwindigkeitsverbindung (z.B. PCIe) mit dem SoC 904 gekoppelt werden kann. Die CPU 918 kann z.B. einen X86-Prozessor enthalten. Die CPU 918 kann zur Ausführung einer Vielfalt von Funktionen verwendet werden, einschließlich zum Beispiel der Beurteilung potenziell inkonsistenter Ergebnisse zwischen ADAS-Sensoren und dem SoC 904 und/oder Überwachung des Status und Zustands der Controller 936 und/oder des Infotainment-SoC 930.

[0129] Das Fahrzeug 900 kann eine GPU 920 (z.B. diskrete GPU oder dGPU) enthalten, die über eine Hochgeschwindigkeitsverbindung (z.B. NVLINK von NVIDIA) mit dem SoC 904 gekoppelt werden kann. Die GPU 920 kann zusätzliche künstliche Intelligenz-Funktionen bereitstellen, z.B. durch die Ausführung redundanter und/oder verschiedener neuronaler Netzwerke, und kann verwendet werden, um neuronale Netzwerke auf Grundlage von Eingaben (z.B. Sensordaten) von Sensoren des Fahrzeugs 900 auszubilden und/oder zu aktualisieren.

[0130] Das Fahrzeug 900 kann außerdem die Netzwerkschnittstelle 924 enthalten, die eine oder mehrere Drahtlosantennen 926 enthalten kann (z.B. eine oder mehrere Funkantennen für verschiedene Kommunikationsprotokolle, wie z.B. eine Mobilfunkantenne, eine Bluetooth-Antenne usw.). Die Netzwerkschnittstelle 924 kann verwendet werden, um

drahtlose Verbindungen über das Internet mit der Cloud (z.B. mit dem Server 978 und/oder anderen Netzwerkgeräten), mit anderen Fahrzeugen und/oder mit Rechenvorrichtungen (z.B. Client-Geräten von Passagieren) zu ermöglichen. Zum Kommunizieren mit anderen Fahrzeugen kann eine direkte Verknüpfung zwischen den zwei Fahrzeugen hergestellt werden und/oder eine indirekte Verknüpfung (z.B. über Netzwerke und über das Internet) hergestellt werden. Direkte Verknüpfungen können unter Verwendung einer Fahrzeug-zu-Fahrzeug-Kommunikationsverknüpfung hergestellt werden. Die Fahrzeug-zu-Fahrzeug-Kommunikationsverbindung kann dem Fahrzeug 900 Informationen über Fahrzeuge in der Nähe des Fahrzeugs 900 liefern (z.B. Fahrzeuge vor, seitlich und/oder hinter dem Fahrzeug 900). Diese Funktion kann Teil einer kooperativen adaptiven Geschwindigkeitsregelfunktion des Fahrzeugs 900 sein.

[0131] Die Netzwerkschnittstelle 924 kann über einen SoC verfügen, der Modulation und Demodulation bietet und es dem Controller 936 ermöglicht, über drahtlose Netzwerke zu kommunizieren. Die Netzwerkschnittstelle 924 kann ein Hochfrequenz-Front-End für die Up-Konvertierung von Basisband zu Hochfrequenz und Down-Konvertierung von Hochfrequenz zu Basisband enthalten. Die Frequenzkonvertierungen können durch hinreichend bekannte Prozesse und/oder unter Verwendung von Überlagerungsverfahren durchgeführt werden. In manchen Beispielen kann die Hochfrequenz-Frontend-Funktionalität durch einen separaten Chip bereitgestellt sein. Die Netzwerkschnittstelle kann eine drahtlose Funktionalität für die Kommunikation über LTE, WCDMA, UMTS, GSM, CDMA2000, Bluetooth, Bluetooth LE, Wi-Fi, Z-Wave, ZigBee, LoRaWAN und/oder andere drahtlose Protokolle beinhalten.

[0132] Das Fahrzeug 900 kann außerdem Datenspeicher 928 enthalten, die auch Off-Chip-Speicher (z.B. Off the-SoC 904) umfassen können. Die Datenspeicher 928 können ein oder mehrere Speicherelemente einschließlich RAM, SRAM, DRAM, VRAM, Flash, Festplatten und/oder andere Komponenten und/oder Geräte, die zumindest ein Bit von Daten speichern können, umfassen.

[0133] Das Fahrzeug 900 kann außerdem GNSS-Sensor(en) 958 enthalten. Der/die GNSS-Sensor(en) 958 (z.B. GPS, unterstützte GPS-Sensoren, Differential-GPS (DGPS)-Sensoren usw.), dient dazu, bei der Kartierung, der Wahrnehmung, der Erstellung von Belegungsrastern und/oder der Wegplanung zu helfen. Es kann eine beliebige Anzahl von GNSS-Sensoren 958 verwendet werden, einschließlich z.B. eines GPS-Geräts, das einen USB-Anschluss mit einer Ethernet-zu-Seriell-Brücke (RS-232) verwendet.

[0134] Das Fahrzeug 900 kann außerdem einen oder mehrere RADAR-Sensoren 960 enthalten. Der RADAR-Sensor 960 kann vom Fahrzeug 900 für die Fernerkennung von Fahrzeugen verwendet werden, selbst bei Dunkelheit und/oder bei Unwetter. Die funktionalen Sicherheitsstufen des RADARS können ASIL B sein. Der RADAR-Sensor 960 kann den CAN und/oder den Bus 902 (z.B. zur Übertragung der vom RADAR-Sensor 960 erzeugten Daten) zur Steuerung und zum Zugriff auf Daten zur Objektverfolgung verwenden, wobei in manchen Beispielen auf Ethernet zugegriffen werden kann, um auf Rohdaten zuzugreifen. Eine große Vielfalt von RADAR-Sensortypen kann verwendet werden. Zum Beispiel können die RADAR-Sensoren 960 uneingeschränkt als Front-, Heck- und Seiten-RADAR verwendet werden. In manchen Beispielen werden Puls-Doppler-RADAR-Sensoren verwendet.

[0135] Die RADAR-Sensoren 960 können verschiedene Konfigurationen beinhalten, z.B. große Reichweite mit engem Sichtfeld, kurze Reichweite mit großem Sichtfeld, kurze Seitenabdeckung usw. In manchen Beispielen kann ein Langstrecken-RADAR für die adaptive Geschwindigkeitsregelung verwendet werden. Die RADAR-Systeme mit großer Reichweite können ein breites Sichtfeld bereitstellen, das durch zwei oder mehr unabhängige Scans realisiert wird, z.B. innerhalb einer Reichweite von 250 m. Die RADAR-Sensoren 960 können bei der Unterscheidung zwischen statischen und sich bewegendenden Objekten helfen und von ADAS-Systemen zur Notbremsunterstützung und Kollisionswarnung verwendet werden. RADAR-Systeme mit großer Reichweite können ein monostatisches multimodales RADAR mit mehreren (z.B. sechs oder mehr) festen RADAR-Antennen und einer Hochgeschwindigkeits-CAN- und FlexRay-Schnittstelle beinhalten. In einem Beispiel mit sechs Antennen können die zentralen vier Antennen ein fokussiertes Strahlmuster erzeugen, das entwickelt wurde, um die Umgebung des Fahrzeugs 900 bei höheren Geschwindigkeiten mit minimaler Beeinträchtigung durch den Verkehr in benachbarten Fahrspuren aufzuzeichnen. Die anderen beiden Antennen können das Sichtfeld erweitern, wodurch es möglich ist, schnell Fahrzeuge zu erkennen, die in die Fahrspur des Fahrzeugs 900 fahren oder diese verlassen.

[0136] Mid-Range-RADAR-Systeme können zum Beispiel eine Reichweite von bis zu 960 m (vorne) oder 80 m (hinten) und ein Sichtfeld von bis zu 42 Grad (vorne) oder 950 Grad (hinten) umfassen. RADAR-Systeme mit kurzer Reichweite können ohne Einschränkung RADAR-Sensoren beinhalten, die für die Installation an beiden Enden des hinteren Stoßfängers ausgelegt sind. Wenn das RADAR-Sensorsystem an beiden Enden des hinteren Stoßfängers installiert ist, kann ein derartiges RADAR-Sensorsystem zwei Strahlen erzeugen, die den toten

Winkel hinter und neben dem Fahrzeug konstant überwachen.

[0137] Short-Range-RADAR-Systeme können in einem ADAS-System zur Erkennung des toten Winkels und/oder zur Spurwechselunterstützung verwendet werden.

[0138] Das Fahrzeug 900 kann außerdem Ultraschallsensoren 962 enthalten. Die Ultraschallsensoren 962, die sich vorne, hinten und/oder an den Seiten des Fahrzeugs 900 befinden können, können zum Einparken und/oder zum Erstellen und Aktualisieren eines Belegungsrasters verwendet werden. Es können eine breite Vielfalt von Ultraschallsensoren 962 und verschiedene Ultraschallsensoren 962 für verschiedene Detektionsbereiche (z.B. 2,5 m, 4 m) verwendet werden. Die Ultraschallsensoren 962 können mit der Funktionssicherheitsstufe ASIL B arbeiten.

[0139] Das Fahrzeug 900 kann LIDAR-Sensoren 964 enthalten. Die LIDAR Sensoren 964 können für die Objekt- und Fußgängererkennung, Notbremsung, Kollisionsvermeidung und/oder andere Funktionen verwendet werden. Die LIDAR-Sensoren 964 können die funktionale Sicherheitsstufe ASIL B sein. In manchen Beispielen kann das Fahrzeug 900 mehrere LIDAR-Sensoren 964 (z.B. zwei, vier, sechs usw.) beinhalten, die Ethernet verwenden können (z.B. zur Bereitstellung von Daten an einen Gigabit-Ethernet-Switch).

[0140] In manchen Beispielen können die LIDAR-Sensoren 964 eine Liste von Objekten und deren Entfernungen für ein 360-Grad-Sichtfeld bereitstellen. Im Handel erhältliche LIDAR-Sensoren 964 können eine angekündigte Reichweite von etwa 900 m, eine Genauigkeit von 2 cm bis 3 cm und zum Beispiel eine Unterstützung für eine 900-Mbit/s-Ethernet-Verbindung aufweisen. In manchen Beispielen können ein oder mehrere nicht vorstehende LIDAR-Sensoren 964 verwendet werden. In solchen Beispielen können die LIDAR-Sensoren 964 als kleines Gerät umgesetzt sein, das in die Front-, Heck-, Seiten- und/oder Ecken des Fahrzeugs 900 eingebettet werden kann. Der/die LIDAR-Sensor(en) 964 können in solchen Beispielen ein horizontales Sichtfeld von bis zu 120 Grad und ein vertikales Sichtfeld von bis zu 35 Grad mit einer Reichweite von 200 m selbst bei Objekten mit niedrigem Reflexionsvermögen bereitstellen. Front-montierte LIDAR-Sensoren 964 können für ein horizontales Sichtfeld zwischen 45 Grad und 135 Grad konfiguriert werden.

[0141] In manchen Beispielen können auch LIDAR-Technologien, wie etwa 3D-Flash-LIDAR, verwendet werden. 3D-Flash-LIDAR verwendet einen Laserblitz als Sendequelle, um die Umgebung des Fahrzeugs bis zu einer Entfernung von ca. 200 m zu beleuchten.

Eine Flash-LIDAR-Einheit beinhaltet einen Rezeptor, der die Laserpuls-Laufzeit und das reflektierte Licht an jedem Pixel aufzeichnet, was wiederum der Reichweite vom Fahrzeug zu den Objekten entspricht. Flash-LIDAR kann ermöglichen, dass mit jedem Laserblitz hochgenaue und verzerrungsfreie Bilder der Umgebung erzeugt werden. In manchen Beispielen können vier LIDAR-Blitzsensoren eingesetzt werden, einer an jeder Seite des Fahrzeugs 900. Verfügbare 3D-Flash-LIDAR-Systeme beinhalten eine Festkörper-3D-Staring-Array-LIDAR-Kamera ohne bewegliche Teile außer einem Lüfter (z.B. eine nicht scannende LIDAR-Vorrichtung). Die Flash-LIDAR-Vorrichtung kann einen 5-Nanosekunden-Laserpuls der Klasse I (augensicher) pro Bild verwenden und kann das reflektierte Laserlicht in Form von den 3D-Reichweitenpunktwolken und gemeinsam registrierten Intensitätsdaten erfassen. Durch die Verwendung von Flash-LIDAR und weil Flash-LIDAR ein Festkörpergerät ohne bewegliche Teile ist, ist der LIDAR-Sensor 964 möglicherweise weniger anfällig für Bewegungsunschärfe, Vibrationen und/oder Stöße.

[0142] Das Fahrzeug kann außerdem IMU-Sensoren 966 enthalten. Die IMU-Sensoren 966 können sich in manchen Beispielen in der Mitte der Hinterachse des Fahrzeugs 900 befinden. Die IMU-Sensoren 966 können z.B. einen oder mehrere Beschleunigungsmesser, Magnetometer, Gyroskope, magnetische Kompassse und/oder andere Sensortypen umfassen. In manchen Beispielen, wie z.B. in sechs-Achsen-Anwendungen, können die IMU-Sensoren 966 Beschleunigungsmesser und Gyroskope enthalten, während in neun-Achsen-Anwendungen die IMU-Sensoren 966 Beschleunigungsmesser, Gyroskope und Magnetometer umfassen können.

[0143] In manchen Ausführungsformen können die IMU-Sensoren 966 als Hochleistungs-Miniatur-GPS-Aided Inertial Navigation System (GPS/INS) umgesetzt sein, das Inertialsensoren mikroelektromechanischer Systeme (micro-electro-mechanical systems - MEMS), einen hochempfindlichen GPS-Empfänger und fortschrittliche Kalman-Filteralgorithmen kombiniert, um Abschätzungen von Position, Geschwindigkeit und Neigung zu liefern. In manchen Beispielen können die IMU-Sensoren 966 es dem Fahrzeug 900 ermöglichen, den Kurs ohne Eingabe eines Magnetsensors zu schätzen, indem die Geschwindigkeitsänderungen vom GPS direkt auf den IMU-Sensoren 966 beobachtet und korreliert werden. In manchen Beispielen können IMU-Sensoren 966 und GNSS-Sensoren 958 in einer einzigen integrierten Einheit kombiniert werden.

[0144] Das Fahrzeug kann mit Mikrofonen 996996 ausgestattet sein, die in und/oder um das Fahrzeug 900 platziert sind. Die Mikrofone 996996 können

unter anderem zur Erkennung und Identifizierung von Rettungsfahrzeugen verwendet werden.

[0145] Das Fahrzeug kann außerdem eine beliebige Anzahl von Kameratypen umfassen, einschließlich Stereokameras 968, Weitwinkelkameras 970, Infrarotkameras 972, Surround-Kameras 974, Long-Range- und Mid-Range-Kameras 998 und/oder andere Kameratypen. Mit den Kameras können Bild-daten rund um eine gesamte Peripherie des Fahrzeugs 900 herum erfasst werden. Die verwendeten Kameratypen hängen von den Ausführungsformen und Anforderungen für das Fahrzeug 900 ab, und jede Kombination von Kameratypen kann verwendet werden, um die erforderliche Abdeckung um das Fahrzeug 900 herum sicherzustellen. Zusätzlich kann die Anzahl der Kameras in Abhängigkeit von der Ausführungsform unterschiedlich sein. Zum Beispiel kann das Fahrzeug sechs Kameras, sieben Kameras, zehn Kameras, zwölf Kameras und/oder eine andere Anzahl von Kameras beinhalten. Die Kameras können zum Beispiel und ohne Einschränkung Gigabit Multimedia Serial Link (GMSL) und/oder Gigabit Ethernet unterstützen. Jede Kamera wird in diesem Dokument in Bezug auf **Fig. 9A** und **Fig. 9B** beschrieben.

[0146] Das Fahrzeug 900 kann außerdem Vibrationssensoren 942 enthalten. Die Vibrationssensoren 942 können Vibrationen von Komponenten des Fahrzeugs, z.B. den Achsen, messen. Zum Beispiel können Änderungen der Vibrationen eine Änderung des Straßenbelags angeben. In einem anderen Beispiel können bei Verwendung von zwei oder mehr Vibrationssensoren 942 die Unterschiede zwischen den Vibrationen zur Bestimmung der Reibung oder des Schlupfes der Fahrspuroberfläche verwendet werden (z.B. wenn der Vibrationsunterschied zwischen einer elektrisch angetriebenen und einer frei drehenden Achse besteht).

[0147] Das Fahrzeug 900 kann ein ADAS-System 938 enthalten. Das ADAS-System 938 kann in manchen Beispielen ein SoC enthalten. Das ADAS-System 938 kann autonome/adaptive/automatische Geschwindigkeitsregelung (autonomous/adaptive/automatic cruise control - ACC), kooperative adaptive Geschwindigkeitsregelung (cooperative adaptive cruise control - CACC), Aufprallwarnung (forward crash warning - FCW), automatische Notbremsung (automatic emergency braking - AEB), Spurverlassenswarnung (lane departure warning - LDW), Spurhalteassistentz (lane keep assist - LKA), Totwinkel-Warner (blind spot warning - BSW), Heckbereichswarnung (rear cross-traffic warning - RCTW), Kollisionswarnsysteme (collision warning system - CWS), Spurzentrierung (lane centering - LC) und/oder andere Funktionen und Funktionen umfassen.

[0148] Die ACC-Systeme können RADAR-Sensoren 960, LIDAR-Sensoren 964 und/oder Kameras verwenden. Die ACC-Systeme können eine ACC in Längsrichtung und/oder eine ACC in Querrichtung beinhalten. Das ACC in Längsrichtung überwacht und steuert den Abstand zum Fahrzeug unmittelbar vor dem Fahrzeug 900 und passt die Fahrzeuggeschwindigkeit automatisch an, um einen sicheren Abstand zu den vorausfahrenden Fahrzeugen einzuhalten. Das laterale ACC führt die Distanzmessung durch und rät dem Fahrzeug 900, bei Bedarf die Fahrspur zu wechseln. Die Quer-ACC ist mit anderen ADAS-Anwendungen, wie zum Beispiel LCA und CWS, verbunden.

[0149] CACC verwendet Informationen von anderen Fahrzeugen, die über die Netzwerkschnittstelle 924 und/oder die Wireless-Antenne(n) 926 von anderen Fahrzeugen über eine Wireless-Verbindung oder indirekt über eine Netzwerkverbindung (z.B. über das Internet) empfangen werden können. Direkte Verknüpfungen können durch eine Fahrzeug-zu-Fahrzeug (V2V)-Kommunikationsverknüpfung bereitgestellt werden, während indirekte Verknüpfungen durch eine Infrastruktur-zu-Fahrzeug (I2V)-Kommunikationsverknüpfung bereitgestellt werden können. Das Kommunikationskonzept V2V informiert in der Regel über die unmittelbar vorausfahrenden Fahrzeuge (z.B. Fahrzeuge unmittelbar vor und auf derselben Fahrspur wie das Fahrzeug 900), während das Kommunikationskonzept I2V Informationen über den weiteren vorausfahrenden Verkehr liefert. CACC-Systeme können entweder eines oder beides von I2V- und V2V-Informationsquellen beinhalten. Angesichts der Informationen über die Fahrzeuge vor dem Fahrzeug 900 könnte CACC zuverlässiger sein und das Potenzial haben, die Verkehrsströme zu verbessern und Staus auf der Straße zu reduzieren.

[0150] FCW-Systeme wurden entwickelt, um den Fahrer auf eine Gefahr aufmerksam zu machen, damit der Fahrer Korrekturmaßnahmen ergreifen kann. FCW-Systeme verwenden eine nach vorne gerichtete Kamera und/oder RADAR-Sensoren 960 gekoppelt an einen dedizierten Prozessor, DSP, FPGA und/oder ASIC, der elektrisch mit dem Fahrer-Feedback gekoppelt ist, z.B. einer Anzeige, einem Lautsprecher und/oder einer vibrierenden Komponente. FCW-Systeme können eine Warnung bereitstellen, z.B. in Form eines Tons, einer optischen Warnung, einer Vibration und/oder eines schnellen Bremsimpulses.

[0151] AEB-Systeme erkennen einen drohenden Aufprall mit einem anderen Fahrzeug oder einem anderen Objekt und können die Bremsen automatisch betätigen, wenn der Fahrer innerhalb eines bestimmten Zeit- oder Entfernungsparameters keine Korrekturmaßnahmen ergreift. AEB-Systeme können nach vorne gerichtete Kameras und/oder

RADAR-Sensoren 960 verwenden, die mit einem dedizierten Prozessor, DSP, FPGA und/oder ASIC gekoppelt sind. Wenn das AEB-System eine Gefahr erkennt, warnt es typischerweise zuerst den Fahrer, um eine korrigierende Handlung durchzuführen, um eine Kollision zu vermeiden, und, wenn der Fahrer keine korrigierende Handlung durchführt, kann das AEB-System automatisch die Bremsen in dem Bestreben betätigen, die Auswirkungen der vorhergesagten Kollision zu verhindern oder zumindest abzuschwächen. AEB-Systeme können Techniken, wie zum Beispiel dynamische Bremsunterstützung und/oder Bremsung aufgrund eines bevorstehenden Zusammenstoßes, beinhalten.

[0152] LDW-Systeme bieten optische, akustische und/oder taktile Warnungen, wie z.B. Lenkrad- oder Sitzvibrationen, um den Fahrer zu warnen, wenn das Fahrzeug 900 die Fahrspurmarkierungen überquert. Ein LDW-System wird nicht aktiviert, wenn der Fahrer ein absichtliches Verlassen der Fahrspur angibt, indem er den Blinker betätigt. LDW-Systeme können nach vorne gerichtete Kameras verwenden, die mit einem dedizierten Prozessor, DSP, FPGA und/oder ASIC gekoppelt sind, die elektrisch mit einer Rückmeldung des Fahrers gekoppelt sind, wie z.B. einer Anzeige, einem Lautsprecher und/oder einer vibrierenden Komponente.

[0153] LKA-Systeme sind eine Variante von LDW-Systemen. LKA-Systeme bieten Lenkeingaben oder Bremsen, um das Fahrzeug 900 zu korrigieren, wenn das Fahrzeug 900 die Fahrspur verlässt.

[0154] BSW-Systeme erkennen und warnen den Fahrer vor Fahrzeugen im toten Winkel eines Automobils. BSW-Systeme können einen optischen, akustischen und/oder taktilen Alarm bereitstellen, um anzugeben, dass Einfädeln in oder Wechseln der Fahrspuren unsicher ist. Das System kann eine zusätzliche Warnung ausgeben, wenn der Fahrer einen Blinker verwendet. BSW-Systeme können rückseitig ausgerichtete Kameras und/oder RADAR-Sensoren 960 verwenden, gekoppelt mit einem dedizierten Prozessor, DSP, FPGA und/oder ASIC, der elektrisch mit dem Fahrer-Feedback gekoppelt ist, z.B. einer Anzeige, einem Lautsprecher und/oder einer vibrierenden Komponente.

[0155] RCTW-Systeme können visuelle, akustische und/oder taktile Benachrichtigungen liefern, wenn ein Objekt außerhalb des Bereichs der Rückfahrkamera erkannt wird, wenn das Fahrzeug 900 rückwärts fährt. manche RCTW-Systeme beinhalten das AEB-System, um sicherzustellen, dass die Fahrzeugbremsen betätigt werden, um einen Zusammenstoß zu vermeiden. RCTW-Systeme können eine oder mehrere nach hinten gerichtete RADAR-Sensoren 960 verwenden, gekoppelt mit einem dedizierten Prozessor, DSP, FPGA und/oder ASIC, der elektrisch

mit dem Fahrer-Feedback gekoppelt ist, z.B. einer Anzeige, einem Lautsprecher und/oder einer vibrierenden Komponente.

[0156] Herkömmliche ADAS-Systeme können zu falsch positiven Ergebnissen neigen, die für den Fahrer störend und ablenkend sein können, aber in der Regel nicht katastrophal sind, da die ADAS-Systeme den Fahrer alarmieren und es dem Fahrer ermöglichen, zu entscheiden, ob eine Sicherheitsbedingung tatsächlich vorliegt, und entsprechend zu handeln. Bei einem autonomen Fahrzeug 900 muss das Fahrzeug 900 selbst jedoch bei widersprüchlichen Ergebnissen entscheiden, ob das Ergebnis eines primären oder eines sekundären Computers (z.B. eines ersten Controllers 936 oder eines zweiten Controllers 936) berücksichtigt werden soll. In manchen Ausführungsformen kann das ADAS-System 938 zum Beispiel ein Backup- und/oder Sekundärcomputer sein, um Wahrnehmungsinformationen für ein Rationalitätsmodul des Backup-Computers bereitzustellen. Der Rationalitätsmonitor des Backup-Computers kann eine redundante, diverse Software auf Hardware-Komponenten ausführen, um Fehler in der Wahrnehmung und bei dynamischen Fahraufgaben zu erkennen. Ausgaben vom ADAS-System 938 können an eine übergeordnete MCU geliefert werden. Wenn die Ausgaben des primären und des sekundären Computers miteinander in Konflikt geraten, muss die Überwachungs-MCU bestimmen, wie der Konflikt beigelegt werden kann, um einen sicheren Betrieb sicherzustellen.

[0157] In manchen Beispielen kann der primäre Computer derart konfiguriert sein, dass er der Überwachungs-MCU eine Konfidenzbewertung bereitstellt, die eine Konfidenz des primären Computers für das gewählte Ergebnis angibt. Falls die Konfidenzbewertung einen Schwellenwert überschreitet, kann diese Überwachungs-MCU der Führung des primären Computers folgen, unabhängig davon, ob dieser sekundäre Computer ein widersprüchliches oder inkonsistentes Ergebnis bereitstellt. Wenn die Konfidenzbewertung den Schwellenwert nicht erreicht und der primäre und der sekundäre Computer unterschiedliche Ergebnisse angeben (z.B. den Widerspruch), kann die Überwachungs-MCU zwischen den Computern bestimmen, um das zweckmäßige Resultat zu bestimmen.

[0158] Die Überwachungs-MCU kann derart konfiguriert sein, dass sie neuronale(s) Netz(e) ausführt, die dafür trainiert und konfiguriert sind, zumindest auf Grundlage von Ausgaben aus dem primären Computer und dem sekundären Computer die Bedingungen zu bestimmen, unter denen der sekundäre Computer Fehlalarme bereitstellt. Folglich können das/die neuronale Netz(e) in der Überwachungs-MCU lernen, wann der Ausgabe eines sekundären Computers vertraut werden kann und wann nicht. Zum Beispiel

können, wenn der sekundäre Computer ein RADAR-basiertes FCW-System ist, neuronale Netz(e) in der Überwachungs-MCU lernen, wann das FCW-System metallische Objekte identifiziert, die tatsächlich keine Gefahren sind, wie etwa ein Abflussgitter oder ein Gullydeckel, das/der einen Alarm auslöst. Ähnlich, wenn der sekundäre Computer ein kamerabasiertes LDW-System ist, kann ein neuronales Netzwerk in der Überwachungs-MCU lernen, die LDW zu überschreiben, wenn Fahrradfahrer oder Fußgänger vorhanden sind und ein Verlassen der Fahrspur tatsächlich das sicherste Manöver ist. In Ausführungsformen, die ein oder mehrere neuronale Netzwerke beinhalten, die auf der Überwachungs-MCU laufen, kann die Überwachungs-MCU zumindest eine DLA oder GPU beinhalten, die für die Ausführung des oder der neuronalen Netzwerke mit zugewiesenem Speicher geeignet ist. In bevorzugten Ausführungsformen kann die Überwachungs-MCU als Bestandteil des SoCs 904 enthalten und/oder einbezogen werden.

[0159] In anderen Beispielen kann das ADAS-System 938 einen sekundären Computer umfassen, der die ADAS-Funktionalität mithilfe herkömmlicher Regeln der Computervision ausführt. Somit kann der sekundäre Computer klassische Regeln des maschinellen Sehens (wenn-dann) verwenden und kann das Vorhandensein eines neuronalen Netzwerks/von neuronalen Netzwerken in der Überwachungs-MCU die Zuverlässigkeit, Sicherheit und Leistung verbessern. Zum Beispiel macht die vielfältige Umsetzung und absichtliche Nicht-Identität das Gesamtsystem fehlertoleranter, insbesondere gegenüber Fehlern, die durch Software (oder Software-Hardware-Schnittstellen)-Funktionalität verursacht werden. Wenn zum Beispiel ein Software-Bug oder -Fehler in der auf dem primären Computer laufenden Software vorliegt und der nicht identische Software-Code, der auf dem sekundären Computer läuft, dasselbe Gesamtergebnis bereitstellt, kann die Überwachungs-MCU eine größere Konfidenz darin haben, dass das Gesamtergebnis korrekt ist und der Bug in der Software oder Hardware auf dem primären Computer keinen wesentlichen Fehler verursacht.

[0160] In manchen Beispielen kann die Ausgabe des ADAS-Systems 938 in den Wahrnehmungs-Block des primären Computers und/oder den dynamischen Fahraufgaben-Block des primären Computers eingespeist werden. Wenn das ADAS-System 938 zum Beispiel eine Aufprallwarnung aufgrund eines unmittelbar vorausliegenden Objekts anzeigt, kann der Wahrnehmungs-Block diese Informationen bei der Identifizierung von Objekten verwenden. In anderen Beispielen kann der sekundäre Computer über sein eigenes neuronales Netzwerk verfügen, das trainiert ist und somit das Risiko von falsch positiven Ergebnissen reduziert, wie hierin beschrieben.

[0161] Das Fahrzeug 900 kann außerdem das Infotainment-SoC 930 (z.B. ein Fahrzeug-Infotainment-System (in-vehicle infotainment - IVI)) enthalten. Obwohl als ein SoC veranschaulicht und beschrieben, kann das Infotainment-System kein SoC sein und kann zwei oder mehr diskrete Komponenten beinhalten. Das Infotainment-SoC 930 kann eine Kombination aus Hardware und Software umfassen, die verwendet werden kann, um dem Fahrzeug 900 Audio (z.B. Musik, einen persönlichen digitalen Assistenten, Navigationsanweisungen, Nachrichten, Radio usw.), Video (z.B. TV, Filme, Streaming usw.), Telefon (z.B. Freisprechfunktion), Netzwerkkonnektivität (z.B. LTE, Wi-Fi, Usw.) und/oder Informationsdienste (z.B. Navigationssysteme, Einparkhilfe hinten, ein Funkdatensystem, fahrzeugbezogene Informationen wie Kraftstoffstand, zurückgelegte Gesamtstrecke, Bremsflüssigkeitsstand, Ölstand, Tür auf/zu, Luftfilterinformationen usw.) bereitzustellen. Zum Beispiel kann das Infotainment-SoC 930 Radios, CD-Player, Navigationssysteme, Videoplayer, USB- und Bluetooth-Konnektivität, Carputer, In-Car-Entertainment, Wi-Fi, Lenkrad-Audiosteuerungen, freihändige Sprachsteuerung, ein Heads-Up-Display (HUD), eine HMI-Anzeige 934, ein Telematikgerät, ein Bedienfeld (z.B. zur Steuerung und/oder Interaktion mit verschiedenen Komponenten, Funktionen und/oder Systemen) und/oder anderen Komponenten umfassen. Das Infotainment-SoC 930 kann außerdem dazu verwendet werden, Informationen (z.B. visuell und/oder hörbar) für Benutzer des Fahrzeugs bereitzustellen, z.B. Informationen aus dem ADAS-System 938, Informationen zum autonomen Fahren wie geplante Fahrzeugmanöver, Bewegungsbahnen, Umgebungsdaten (z.B. Kreuzungsinformationen, Fahrzeuginformationen, Straßeninformationen usw.) und/oder andere Informationen.

[0162] Das Infotainment-SoC 930 kann GPU-Funktionalität enthalten. Das Infotainment-SoC 930 kann über den Bus 902 (z.B. CAN-Bus, Ethernet usw.) mit anderen Geräten, Systemen und/oder Komponenten des Fahrzeugs 900 kommunizieren. In manchen Beispielen kann das Infotainment-SoC 930 mit einer übergeordneten MCU gekoppelt werden, sodass die GPU des Infotainment-Systems manche Selbstfahrfunktionen ausführen kann, falls der primäre Controller 936 (z.B. der primäre und/oder Backup-Computer des Fahrzeugs 900) ausfällt. In einem solchen Beispiel kann das Infotainment-SoC 930 das Fahrzeug 900 in den sicheren Stopp-Modus versetzen, wie hier beschrieben.

[0163] Das Fahrzeug 900 kann außerdem ein Kombiinstrument 932 (z.B. ein digitales Armaturenbrett, ein elektronisches Kombiinstrument, eine digitale Instrumententafel usw.) enthalten. Das Kombiinstrument 932 kann einen Controller und/oder Supercomputer (z.B. einen diskreten Controller oder Super-

computer) enthalten. Das Kombiinstrument 932 kann eine Reihe von Instrumenten wie Tachometer, Kraftstoffstand, Öldruck, Drehzahlmesser, Kilometerzähler, Blinker, Schaltstellungsanzeige, Sicherheitsgurtwarnleuchte(n), Feststellbremse-Warnleuchte(n), Motorstörungsanzeige(n), Airbag (SRS)-Systeminformationen, Beleuchtungssteuerungen, Bedienelemente des Sicherheitssystems, Navigationsinformationen usw. umfassen. In manchen Beispielen können Informationen angezeigt und/oder zwischen dem Infotainment-SoC 930 und dem Kombiinstrument 932 geteilt werden. Mit anderen Worten kann das Kombiinstrument 932 als Teil des Infotainment-SoC 930 oder umgekehrt aufgenommen werden.

[0164] Fig. 9D ein Systemdiagramm für die Kommunikation zwischen Cloudbasierten Servern und dem beispielhaften autonomen Fahrzeug 900 aus Fig. 9A, gemäß manchen Ausführungsformen der vorliegenden Offenbarung. Das System 876 kann Server 978, Netzwerke 990 und Fahrzeuge einschließlich Fahrzeug 900 umfassen. Die Server 978 können eine Vielzahl von GPUs 984(A)-984(H) (zusammenfassend als GPUs 984 bezeichnet), PCIe-Switches 982(A)-982(H) (zusammenfassend als PCIe-Switches 982 bezeichnet) und/oder CPUs 980(A)-980(B) (zusammenfassend als CPUs 980 bezeichnet) enthalten. Die GPUs 984, die CPUs 980 und die PCIe-Switches können mit Hochgeschwindigkeitsverbindungen wie zum Beispiel den von NVIDIA entwickelten NVLink-Schnittstellen 988 und/oder PCIe Connections 986 verbunden werden. In manchen Beispielen werden die GPUs 984 über NVLink und/oder NVSwitch SoC angeschlossen und die GPUs 984 und die PCIe-Switches 982 über PCIe-Verbindungen. Obwohl acht GPUs 984, zwei CPUs 980 und zwei PCIe-Switches abgebildet sind, soll dies nicht als einschränkend verstanden werden. Je nach Ausführungsform kann jeder Server 978 eine beliebige Anzahl von GPUs 984, CPUs 980 und/oder PCIe-Switches enthalten. Zum Beispiel können die Server 978 jeweils acht, sechzehn, zwei- und dreißig und/oder mehr GPUs 984 enthalten.

[0165] Die Server 978 können über die Netzwerke 990 und von den Fahrzeugen Bilddaten empfangen, die für Bilder stehen, die unerwartete oder veränderte Straßenbedingungen zeigen, wie z.B. kürzlich begonnene Straßenarbeiten. Die Server 978 können, über die Netzwerke 990 und an die Fahrzeuge, neuronale Netzwerke 992, aktualisierte neuronale Netzwerke 992 und/oder Karteninformationen 994 übertragen, einschließlich Informationen über den Verkehr und die Straßenbedingungen. Die Aktualisierungen der Karteninformationen 994 können Aktualisierungen für die HD-Karte 922 enthalten, wie z.B. Informationen zu Baustellen, Schlaglöchern, Umwegen, Überschwemmungen und/oder anderen Hindernissen. In manchen Beispielen können die

neuronalen Netzwerke 992, die aktualisierten neuronalen Netzwerke 992 und/oder die Karteninformationen 994 aus neuem Training und/oder neuen Erfahrungen resultieren, dargestellt in Daten aus einer beliebigen Anzahl von Fahrzeugen in der Umgebung, und/oder basierend auf Training, das in einem Rechenzentrum durchgeführt wird (z.B. unter Verwendung der Server 978 und/oder anderer Server).

[0166] Der/die Server 978 können verwendet werden, um Maschinenlernmodelle (z.B. neuronale Netzwerke) basierend auf Trainingsdaten zu trainieren. Die Trainingsdaten von Fahrzeugen können erzeugt werden und/oder können sie in einer Simulation (z.B. unter Verwendung einer Spiele-Engine) erzeugt werden. In manchen Beispielen werden die Trainingsdaten mit Tags versehen (z.B. wenn das neuronale Netzwerk von überwachtem Lernen profitiert) und/oder einer anderen Vorverarbeitung unterzogen, während in anderen Beispielen die Trainingsdaten nicht mit Tags versehen und/oder vorverarbeitet werden (z.B. wenn das neuronale Netzwerk kein überwachtes Lernen benötigt). Das Training kann nach einer oder mehreren Klassen von maschinellen Lerntechniken erfolgen, einschließlich, aber nicht beschränkt auf Klassen wie: überwachtes Training, halbüberwachtes Training, unüberwachtes Training, Selbstlernen, Verstärkungslernen, föderiertes Lernen, Transferlernen, Merkmalslernen (einschließlich Hauptkomponenten- und Clusteranalysen), multilinelles Unterraumlernen, vielfältiges Lernen, Repräsentationslernen (einschließlich Ersatzwörterbuchlernen), regelbasiertes maschinelles Lernen, Anomalieerkennung und alle Varianten oder Kombinationen davon. Nach der Trainieren der Maschinenlernmodelle können die Maschinenlernmodelle von den Fahrzeugen verwendet werden (z.B. über das Netzwerk(e) 990 an die Fahrzeuge übertragen) und/oder die Maschinenlernmodelle können von dem Server(n) 978 zur Fernüberwachung der Fahrzeuge verwendet werden.

[0167] In manchen Beispielen kann der Server 978 Daten von den Fahrzeugen empfangen und die Daten auf aktuelle neuronale Echtzeit-Netzwerke anwenden, um intelligente Echtzeit-Inferenzierungen zu ermöglichen. Die Server 978 können Deep-Learning-Supercomputer und/oder dedizierte KI-Computer mit GPU(s) 984 umfassen, wie z.B. DGX- und DGX-Stationsmaschinen, die von NVIDIA entwickelt wurden. In manchen Beispielen können die Server 978 jedoch eine Deep-Learning-Infrastruktur enthalten, die nur CPUbetriebene Rechenzentren verwendet.

[0168] Die Deep-Learning-Infrastruktur der Server 978 kann eine schnelle Echtzeit-Inferenzierung ermöglichen und diese Funktion nutzen, um den Zustand der Prozessoren, der Software und/oder

der zugehörigen Hardware im Fahrzeug 900 zu bewerten und zu überprüfen. Zum Beispiel kann die Deep-Learning-Infrastruktur regelmäßige Aktualisierungen vom Fahrzeug 900 erhalten, wie z.B. eine Abfolge von Bildern und/oder Objekten, die das Fahrzeug 900 in dieser Abfolge von Bildern lokalisiert hat (z.B. durch Computervision und/oder andere Techniken zur Klassifizierung von Machine-Learning-Objekten). Die Deep-Learning-Infrastruktur kann ein eigenes neuronales Netzwerk betreiben, um die Objekte zu identifizieren und sie mit den vom Fahrzeug 900 identifizierten Objekten zu vergleichen. Wenn die Ergebnisse nicht übereinstimmen und die Infrastruktur zu dem Schluss kommt, dass die KI im Fahrzeug 900 defekt ist, kann der Server 978 ein Signal an das Fahrzeug 900 senden, das einen ausfallsicheren Computer des Fahrzeugs 900 anweist, die Kontrolle zu übernehmen, die Passagiere zu benachrichtigen und ein sicheres Einparkmanöver durchzuführen.

[0169] Für die Inferenzierung können die Server 978 die GPU(s) 984 und einen oder mehrere programmierbare Inferenzbeschleuniger (z.B. NVIDIA TensorRT) enthalten. Die Kombination von GPU-angetriebenen Servern und Inferenzbeschleunigung kann eine Reaktionsfähigkeit in Echtzeit ermöglichen. In anderen Beispielen, wenn z.B. die Performance weniger kritisch ist, können von CPUs, FPGAs und anderen Prozessoren angetriebene Server für die Inferenzierung verwendet werden.

BEISPIELHAFTE RECHENVORRICHTUNG

[0170] Fig. 10 ist ein Blockdiagramm einer beispielhaften Rechenvorrichtung(en) 1000, die zur Verwendung bei der Umsetzung mancher Ausführungsformen der vorliegenden Offenbarung geeignet ist. Die Rechenvorrichtung 1000 kann ein Verschaltungssystem 1002 beinhalten, das die folgenden Vorrichtungen direkt oder indirekt koppelt: Speicher 1004, eine oder mehrere Zentraleinheiten (CPUs) 1006, eine oder mehrere Grafikverarbeitungseinheiten (GPUs) 1008, eine Kommunikationsschnittstelle 1121010, Eingabe-/Ausgabe-(E/A)-Ports 1012, Eingabe-/Ausgabekomponenten 1014, eine Stromversorgung 1016, eine oder mehrere Darstellungskomponenten 1018 (z.B. Anzeige(n)) und eine oder mehrere Logikeinheiten 1020. In zumindest einer Ausführungsform kann/können die Rechenvorrichtung(en) 1000 eine oder mehrere virtuelle Maschinen (VM) umfassen und/oder jede ihrer Komponenten kann aus virtuellen Komponenten bestehen (z.B. virtuelle Hardwarekomponenten). Nicht einschränkende Beispiele können eine oder mehrere GPUs 1008 eine oder mehrere vGPUs umfassen, eine oder mehrere CPUs 1006 können eine oder mehrere vCPUs umfassen und/oder eine oder mehrere Logikeinheiten 1020 können eine oder mehrere virtuelle Logikeinheiten umfassen. Die Rechenvorrichtung(en)

1000 kann (können) somit diskrete Komponenten (z.B. eine vollständige, der Rechenvorrichtung 1000 zugeordnete GPU), virtuelle Komponenten (z.B. einen Teil einer der Rechenvorrichtung 1000 zugeordneten GPU) oder eine Kombination davon beinhalten.

[0171] Auch wenn die verschiedenen Blöcke aus **Fig. 10** als über das Verschaltungssystem 1002 mit Leitungen verbunden gezeigt sind, soll dies nicht einschränkend sein und dient nur der Klarheit. In manchen Ausführungsformen kann zum Beispiel eine Darstellungskomponente 1018, wie z.B. ein Anzeigegerät, als E/A-Komponente 1014 betrachtet werden (z.B. wenn es sich bei der Anzeige um einen Touchscreen handelt). Als weiteres Beispiel können die CPUs 1006 und/oder GPUs 1008 Speicher enthalten (z.B. kann der Speicher 1004 neben dem Speicher der GPUs 1008, der CPUs 1006 und/oder anderer Komponenten repräsentativ für ein Speichergerät sein). Mit anderen Worten ist die Rechenvorrichtung aus **Fig. 10** lediglich illustrativ. Es wird nicht zwischen Kategorien wie „Workstation“, „Server“, „Laptop“, „Desktop“, „Tablet“, „Client-Gerät“, „Mobilgerät“, „Handheld-Gerät“, „Spielkonsole“, „elektronisches Steuergerät (electronic control unit - ECU)“, „Virtual-Reality-System“ und/oder anderen Geräte- oder Systemtypen unterschieden, da alle im Rahmen der Rechenvorrichtung aus **Fig. 10** erwogen werden.

[0172] Das Verschaltungssystem 1002 kann eine oder mehrere Verbindungen oder Busse darstellen, wie zum Beispiel einen Adressbus, einen Datenbus, einen Steuerbus oder eine Kombination davon. Das Verschaltungssystem 1002 kann einen oder mehrere Bus- oder Verbindungstypen beinhalten, wie etwa einen Bus mit Industriestandardarchitektur (ISA), einen Bus mit erweiterter Industriestandardarchitektur (EISA), einen Bus der Video Electronic Standards Association (VESA), einen Bus für Verschaltung einer Periphärkomponente (PCI), einen Bus für Expressverschaltung einer Periphärkomponente (PCIe) und/oder eine andere Art von Bus oder Verbindung. In manchen Ausführungsformen gibt es direkte Verbindungen zwischen Komponenten. Als ein Beispiel kann die CPU 1006 direkt mit dem Speicher 1004 verbunden sein. Ferner kann die CPU 1006 direkt mit der GPU 1008 verbunden sein. Wo eine direkte oder Punkt-zu-Punkt-Verbindung zwischen Komponenten besteht, kann das Verschaltungssystem 1002 eine PCIe-Verbindung beinhalten, um die Verbindung auszuführen. In diesen Beispielen muss kein PCI-Bus in der Rechenvorrichtung 1000 beinhaltet sein.

[0173] Der Speicher 1004 kann eine Vielfalt von computerlesbaren Medien beinhalten. Die computerlesbaren Medien können alle verfügbaren Medien sein, auf die die Rechenvorrichtung 1000 zugreifen

kann. Die computerlesbaren Medien können sowohl flüchtige als auch nichtflüchtige Medien sowie entfernbare und nicht entfernbare Medien beinhalten. Beispielfhaft und nicht einschränkend können die computerlesbaren Medien Computerspeichermedien und Kommunikationsmedien umfassen.

[0174] Das Computerspeichermedium kann sowohl flüchtige als auch nichtflüchtige Medien und/oder wechselbare und nicht wechselbare Medien enthalten, die in jedem Verfahren oder jeder Technologie zur Speicherung von Informationen wie computerlesbaren Anweisungen, Datenstrukturen, Programmmodulen und/oder andere Datentypen umgesetzt sind. Im Speicher 1004 können zum Beispiel computerlesbare Anweisungen gespeichert werden (z.B. die ein oder mehrere Programme und/oder Programmelemente darstellen, z.B. ein Betriebssystem). Die Speichermedien für Computer können RAM, ROM, EEPROM, Flash-Speicher oder andere Speichertechnologien, CD-ROM, Digital Versatile Disks (DVD) oder andere optische Datenträger, Magnetkassetten, Magnetband, Magnetplattenspeicher oder andere magnetische Speichergeräte, oder jedes andere Medium, das zum Speichern der gewünschten Informationen verwendet werden kann und auf das die Rechenvorrichtung 1000 zugreifen kann, beinhalten, sind darauf aber nicht beschränkt. Im hierin verwendeten Sinne umfassen Computerspeichermedien keine Signale an sich.

[0175] Die Computerspeichermedien können computerlesbare Anweisungen, Datenstrukturen, Programmmodule und/oder andere Datentypen in einem modulierten Datensignal wie etwa einer Trägerwelle oder einem anderen Transportmechanismus verkörpern und beinhalten beliebige Informationsliefermedien. Der Begriff „moduliertes Datensignal“ kann ein Signal bezeichnen, das eine oder mehrere seiner Eigenschaften auf solch eine Weise verändert aufweist, dass Informationen in dem Signal kodiert werden. Zum Beispiel, und nicht als Einschränkung, können Computerspeichermedien verkabelte Medien beinhalten, wie zum Beispiel ein verkabeltes Netzwerk oder eine drahtgebundene Verbindung, und drahtlose Medien, wie zum Beispiel akustische, RF, infrarote und andere drahtlose Medien. Kombinationen aller Vorstehenden sollen ebenfalls im Umfang computerlesbarer Medien eingeschlossen sein.

[0176] Die CPU(s) 1006 können konfiguriert sein, um zumindest manche der computerlesbaren Anweisungen auszuführen, um eine oder mehrere Komponenten der Rechenvorrichtung 1000 zu steuern, um eines/einen oder mehrere der Verfahren und/oder Prozesse, die hierin beschrieben sind, auszuführen. Die CPU(s) 1006 können jeweils einen oder mehrere Kerne (z.B. einen, zwei, vier, acht, achtundzwanzig, zweiundsiebzig usw.) umfassen, die in der Lage sind,

eine Vielzahl von Software-Threads gleichzeitig zu verarbeiten. Die CPU(s) 1006 können jede beliebige Art von Prozessor beinhalten und je nach Art der umgesetzten Rechenvorrichtung 1000 verschiedene Typen von Prozessoren beinhalten (z.B. Prozessoren mit weniger Kernen für Mobilgeräte und Prozessoren mit mehr Kernen für Server). Zum Beispiel kann der Prozessor in Abhängigkeit von der Art der Rechenvorrichtung 1000 ein Advanced-RISC-Machines(ARM)-Prozessor sein, der unter Verwendung von Reduced Instruction Set Computing (RISC) umgesetzt ist, oder ein x86-Prozessor, der unter Verwendung von Complex Instruction Set Computing (CISC) umgesetzt ist. Die Rechenvorrichtung 1000 kann zusätzlich zu einem oder mehreren Mikroprozessoren oder zusätzlichen Koprozessoren, wie z.B. mathematischen Koprozessoren, eine oder mehrere CPUs 1006 beinhalten.

[0177] Zusätzlich oder alternativ zu den CPU(s) 1006 können die GPU(s) 1008 dazu konfiguriert sein, zumindest manche der computerlesbaren Anweisungen auszuführen, um eine oder mehrere Komponenten der Rechenvorrichtung 1000 zu steuern, um eines/einen oder mehrere der Verfahren und/oder Prozesse, die hierin beschrieben sind, auszuführen. Eine oder mehrere der GPU(s) 1008 können eine integrierte GPU sein (z.B. mit einer oder mehreren der CPU(s) 1006 und/oder eine oder mehrere der GPU(s) 1008 können eine diskrete GPU sein. In Ausführungsformen können eine oder mehrere der GPU(s) 1008 ein Coprozessor einer oder mehrerer CPU(s) 1006 sein. Die GPU(s) 1008 können durch die Rechenvorrichtung 1000 verwendet werden, um Grafiken (z.B. 3D-Grafiken) zu rendern oder Universalberechnungen durchzuführen. Zum Beispiel können die GPU(s) 1008 für Universalberechnungen auf GPUs (GPGPU) verwendet werden. Die GPUs 1008 können Hunderte oder Tausende von Kernen umfassen, die in der Lage sind, Hunderte oder Tausende von Software-Threads gleichzeitig zu verarbeiten. Die GPUs 1008 können als Reaktion auf Rendering-Befehle (z.B. Rendering-Befehle von den CPU(s) 1006, die über eine Host-Schnittstelle empfangen wurden) Pixeldaten für Ausgabebilder erzeugen. Die GPU(s) 1008 können Grafikspeicher beinhalten, wie etwa Anzeigespeicher, um Pixeldaten oder andere geeignete Daten zu speichern, wie etwa GPGPU-Daten. Der Anzeigespeicher kann als Teil des Speichers 1004 beinhaltet sein. Die GPU(s) 1008 können zwei oder mehrere GPUs beinhalten, die parallel arbeiten (z.B. über eine Verbindung). Die Verbindung kann die GPUs direkt verbinden (z.B. unter Verwendung von NVLINK) oder kann die GPUs über ein Switch verbinden (z.B. unter Verwendung von NVSwitch). Wenn sie miteinander kombiniert werden, kann jede GPU 1008 Pixeldaten oder GPGPU-Daten für verschiedene Abschnitte einer Ausgabe oder für verschiedene Ausgaben (z.B. eine erste GPU für ein erstes Bild und eine zweite

GPU für ein zweites Bild) erzeugen. Jede GPU kann ihren eigenen Speicher beinhalten oder kann Speicher mit anderen GPUs teilen.

[0178] Zusätzlich oder alternativ zu den CPU(s) 1006 und/oder den GPU(s) 1008 kann/können die Logikeinheit(en) 1020 dazu konfiguriert sein, zumindest manche der computerlesbaren Anweisungen auszuführen, um eine oder mehrere Komponenten der Rechenvorrichtung 1000 zu steuern, um eines/einen oder mehrere der Verfahren und/oder Prozesse, die hierin beschrieben sind, auszuführen. In Ausführungsformen können die CPU(s) 1006, die GPU(s) 1008 und/oder die Logikeinheit(en) 1020 einzeln oder gemeinsam eine beliebige Kombination der Verfahren, Prozesse und/oder Teile davon ausführen. Eine oder mehrere der Logikeinheiten 1020 kann/können Teil von und/oder integriert in eine oder mehrere der CPU(s) 1006 und/oder der GPU(s) 1008 sein und/oder eine oder mehrere der Logikeinheiten 1020 kann/können diskrete Komponenten oder anderweitig extern zu der/den CPU(s) 1006 und/oder der/den GPU(s) 1008 sein. In Ausführungsformen können eine oder mehrere der Logikeinheiten 1020 ein Coprozessor einer oder mehrerer der CPU(s) 1006 und/oder einer oder mehrerer der GPU(s) 1008 sein.

[0179] Beispiele der Logikeinheit(en) 1020 beinhalten einen oder mehrere Verarbeitungskerne und/oder Komponenten davon, wie etwa Datenverarbeitungseinheiten (Data Processing Units - DPUs), Tensorkerne (Tensor Cores - TC), Tensor-Verarbeitungseinheiten (Tensor Processing Unit - TPU), visuelle Pixelkerne (Pixel Visual Cores - PVC), Bildverarbeitungseinheiten (Vision Processing Unit - VPU), Grafikverarbeitungscluster (Graphics Processing Cluster - GPC), Texturverarbeitungscluster (Texture Processing Cluster - TPC), Streaming-Multiprozessoren (SM), Baumdurchquerungseinheiten (Tree Traversal Unit - TTU), Beschleuniger für künstliche Intelligenz (Artificial Intelligence Accelerator - AIA), Deep-Learning-Beschleuniger (Deep Learning Accelerator - DLA), arithmetische Logikeinheiten (ALU), anwendungsspezifische integrierte Schaltungen (ASIC), Gleitkommaeinheiten (Floating Point Unit - FPU), Eingabe/Ausgabe(E/A)-Elemente, Elemente für Verschaltung von Periphärkomponenten (PCI) oder Expressverschaltung von Periphärkomponenten (peripheral component interconnect express - PCIe) und/oder dergleichen.

[0180] Die Kommunikationsschnittstelle 11121010 kann einen oder mehrere Empfänger, Sender und/oder Transceiver beinhalten, die es der Rechenvorrichtung 1000 ermöglichen, mit anderen Rechenvorrichtungen über ein elektronisches Kommunikationsnetzwerk, einschließlich drahtgebundener und/oder drahtloser Kommunikation, zu kommunizieren. Die Kommunikationsschnittstelle

11121010 kann Komponenten und Funktionalität beinhalten, um eine Kommunikation über eine Anzahl unterschiedlicher Netzwerke zu ermöglichen, wie etwa drahtlose Netzwerke (z.B. Wi-Fi, Z-Wave, Bluetooth, Bluetooth LE, ZigBee usw.), drahtgebundene Netzwerke (z.B. Kommunikation über Ethernet oder InfiniBand), Weiterverkehrsnetzwerke mit geringer Leistung (z.B. LoRaWAN, SigFox usw.) und/oder das Internet. In einer oder mehreren Ausführungsformen können die Logikeinheit(en) 1020 und/oder die Kommunikationsschnittstelle 11121010 eine oder mehrere Datenverarbeitungseinheiten (DPUs) beinhalten, um über ein Netzwerk und/oder durch das Verschaltungssystem 1002 empfangene Daten direkt an eine oder mehrere GPU(s) 1008 (z.B. einen Speicher davon) zu übertragen.

[0181] Die E/A-Ports 1012 können ermöglichen, dass die Rechenvorrichtung 1000 logisch mit anderen Vorrichtungen gekoppelt wird, einschließlich der E/A-Komponenten 1014, der Darstellungskomponente(n) 1018 und/oder anderen Komponenten, von denen manche in die Rechenvorrichtung 1000 eingebaut (z.B. integriert) sein können. Die dargestellten E/A-Komponenten 1014 beinhalten ein Mikrofon, eine Maus, eine Tastatur, einen Joystick, ein Gamepad, einen Gamecontroller, eine Satellitenschüssel, einen Scanner, einen Drucker, ein drahtloses Gerät usw. Die E/A-Komponenten 1014 können eine natürliche Benutzeroberfläche (natural user interface - NUI) bereitstellen, die Luftgesten, Stimme oder andere physiologische Eingaben verarbeitet, die von einem Benutzer erzeugt werden. In manchen Fällen können Eingaben zur weiteren Verarbeitung an ein geeignetes Netzwerkelement übertragen werden. Eine NUI kann eine beliebige Kombination aus Spracherkennung, Tasterkennung, Gesichtserkennung, biometrischer Erkennung, Gestenerkennung sowohl auf dem Bildschirm als auch in der Nähe des Bildschirms, Luftgesten, Kopf- und Augenverfolgung und Touch-Erkennung (wie unten näher beschrieben) umsetzen, die mit einer Anzeige der Rechenvorrichtung 1000 verbunden ist. Die Rechenvorrichtung 1000 kann Tiefenkameras beinhalten, wie stereoskopische Kamerasysteme, Infrarotkamerasysteme, RGB-Kamerasysteme, Touchscreen-Technologie und Kombinationen davon, zur Gestendetektion und -erkennung. Darüber hinaus kann die Rechenvorrichtung 1000 Beschleunigungsmesser oder Gyroskope (z.B. als Teil einer Trägheitsmeseinheit (IMU)) beinhalten, die eine Bewegungserkennung ermöglichen. In manchen Beispielen kann die Ausgabe der Beschleunigungsmesser oder Gyroskope von der Rechenvorrichtung 1000 verwendet werden, um immersive Augmented Reality oder Virtual Reality zu rendern.

[0182] Die Stromversorgung 1016 kann eine fest verdrahtete Stromversorgung, eine Batteriestromversorgung oder eine Kombination davon beinhalten.

Die Stromversorgung 1016 kann die Rechenvorrichtung 1000 mit Strom versorgen, damit die Komponenten der Rechenvorrichtung 1000 funktionieren können.

[0183] Die Darstellungskomponenten 1018 können eine Anzeige (z.B. einen Monitor, einen Touchscreen, einen Fernsehbildschirm, ein Heads-Up-Display (HUD), andere Anzeigetypen oder eine Kombination davon), Lautsprecher und/oder andere Darstellungskomponenten beinhalten. Die Darstellungskomponenten 1018 können Daten von anderen Komponenten (z.B. den GPU(s) 1008, den CPU(s) 1006, DPUs usw.) empfangen und die Daten ausgeben (z.B. als Bild, Video, Ton usw.).

BEISPIELHAFTES RECHENZENTRUM

[0184] Fig. 11 veranschaulicht ein beispielhaftes Rechenzentrum 1100, das in zumindest einer Ausführungsform der vorliegenden Offenbarung verwendet werden kann. Das Rechenzentrum 1100 kann eine Rechenzentrums-Infrastrukturschicht 1110, eine Framework-Schicht 1120, eine Softwareschicht 1130 und/oder eine Anwendungsschicht 1140 beinhalten.

[0185] Wie in Fig. 11 gezeigt, kann die Rechenzentrums-Infrastrukturschicht 1110 einen Ressourcen-Orchestrator 1112, gruppierte Rechenressourcen 1114 und Knoten-Rechenressourcen („Knoten-C.R.s“) 1116(1)-1116(N) beinhalten, wobei „N“ eine beliebige positive ganze Zahl darstellt. In zumindest einer Ausführungsform können die Knoten-C.R.s 1116(1)-1116(N) eine beliebige Anzahl von Zentraleinheiten (CPUs) oder andere Prozessoren (einschließlich DPUs, Beschleuniger, feldprogrammierbare Gate-Arrays (FPGAs), Grafikprozessoren oder Grafikverarbeitungseinheiten (GPUs) usw.), Speichervorrichtungen (z.B. dynamischer Nur-Lese-Speicher), Datenspeichervorrichtungen (z.B. Festkörper- oder Festplattenlaufwerke), Netzwerk-Ein-/Ausgabevorrichtungen (NW E/A), Netzwerk-Switches, virtuellen Maschinen (VMs), Leistungsmodul und/oder Kühlmodule usw. beinhalten, ohne darauf beschränkt zu sein. In zumindest einer Ausführungsform können ein oder mehrere Knoten-C.R.s unter den Knoten-C.R.s 1116(1)-1116(N) einem Server entsprechen, der eine oder mehrere der vorstehend genannten Rechenressourcen aufweist. Darüber hinaus können die Knoten-C.R.s 1116(1)-1116(N) in manchen Ausführungsformen eine oder mehrere virtuelle Komponenten beinhalten, wie etwa vGPUs, vCPUs und/oder dergleichen, und/oder können einer oder mehreren der Knoten-C.R.s 1116(1)-1116(N) einer virtuellen Maschine (VM) entsprechen.

[0186] In zumindest einer Ausführungsform können die gruppierten Rechenressourcen 1114 separate Gruppierungen von Knoten-C.R.s 1116 beinhalten,

die in einem oder mehreren Racks (nicht gezeigt) oder vielen Racks untergebracht sind, die in Rechenzentren an verschiedenen geografischen Standorten untergebracht sind (ebenfalls nicht gezeigt). Separate Gruppierungen von Knoten-C.R.s 1116 innerhalb der gruppierten Rechenressourcen 1114 können gruppierte Rechen-, Netzwerk-, Arbeitsspeicher- oder Datenspeicherressourcen beinhalten, die zur Unterstützung einer oder mehrerer Arbeitslasten konfiguriert oder zugewiesen sein können. In zumindest einer Ausführungsform können mehrere Knoten-C.R.s 1116, die CPUs, GPUs, DPUs und/oder Prozessoren beinhalten, in einem oder mehreren Racks gruppiert sein, um Rechenressourcen bereitzustellen, um eine oder mehrere Arbeitslasten zu unterstützen. Das eine oder die mehreren Racks können auch eine beliebige Anzahl von Leistungsmodulen, Kühlmodulen und/oder Netzwerk-Switches in beliebiger Kombination beinhalten.

[0187] Der Ressourcen-Orchestrator 1112 kann einen oder mehrere Knoten-C.R.s 1116(1)-1116(N) und/oder gruppierte Rechenressourcen 1114 konfigurieren oder anderweitig steuern. In zumindest einer Ausführungsform kann der Ressourcen-Orchestrator 1112 eine Verwaltungseinheit für Software-Design-Infrastruktur (SDI) für das Rechenzentrum 1100 beinhalten. Der Ressourcen-Orchestrator 1112 kann Hardware, Software oder eine Kombination davon beinhalten.

[0188] In zumindest einer Ausführungsform, wie in **Fig. 11** gezeigt, kann die Framework-Schicht 1120 einen Aufgabenplaner 1132, einen Konfigurationsmanager 1134, einen Ressourcenmanager 1136 und/oder ein verteiltes Dateisystem 1138 beinhalten. Die Framework-Schicht 1120 kann einen Rahmen zum Unterstützen von Software 1132 der Software-Schicht 1130 und/oder einer oder mehreren Anwendung(en) 1142 der Anwendungsschicht 1140 beinhalten. Die Software 1132 oder die Anwendung(en) 1142 können jeweils webbasierte Dienstsoftware oder -anwendungen beinhalten, wie etwa diejenigen, die von Amazon Web Services, Google Cloud und Microsoft Azure bereitgestellt werden. Die Framework-Schicht 1120 kann eine Art von freiem und Open-Source-Software-Webanwendungs-Framework sein, ohne darauf beschränkt zu sein, wie etwa Apache Spark™ (im Folgenden „Spark“), welches das verteilte Dateisystem 1138 für umfangreiche Datenverarbeitungen (z.B. „Big Data“) nutzen kann. In zumindest einer Ausführungsform kann der Aufgabenplaner 1132 einen Spark-Treiber beinhalten, um die Planung von Arbeitslasten zu erleichtern, die durch verschiedene Schichten des Rechenzentrums 1100 unterstützt werden. Der Konfigurationsmanager 1134 kann in der Lage sein, unterschiedliche Schichten zu konfigurieren, wie etwa die Software-schicht 1130 und die Framework-Schicht 1120, was Spark und das verteilte Dateisystem 1138 zum

Unterstützen einer umfangreichen Datenverarbeitung beinhaltet. Der Ressourcenmanager 1136 kann in der Lage sein, geclusterte oder gruppierte Rechenressourcen zu verwalten, die zur Unterstützung des verteilten Dateisystems 1138 und des Aufgabenplaners 1132 zugeordnet oder zugewiesen sind. In zumindest einer Ausführungsform können geclusterte oder gruppierte Rechenressourcen die gruppierte Rechenressource 1114 auf der Rechenzentrumsinfrastrukturschicht 1110 beinhalten. Der Ressourcenmanager 1136 und der Ressourcen-Orchestrator 1112 können sich aufeinander abstimmen, um diese zugeordneten oder zugewiesenen Rechenressourcen zu verwalten.

[0189] In zumindest einer Ausführungsform kann die in der Software-Schicht 1130 enthaltene Software 1132 Software beinhalten, die zumindest von Abschnitten der Knoten-C.R.s 1116(1)-1116(N), den gruppierten Rechenressourcen 1114 und/oder dem verteilten Dateisystem 1138 der Framework-Schicht 1120 verwendet wird. Zu einem oder mehreren Typen von Software können Software zum Durchsuchen von Internet-Webseiten, Software zum Scannen von E-Mails auf Viren, Datenbank-Software und Software für Streaming-Videoinhalte gehören, ohne darauf beschränkt zu sein.

[0190] In zumindest einer Ausführungsform können die in der Anwendungsschicht 1140 enthaltenen Anwendung(en) 1142 einen oder mehrere Typen von Anwendungen beinhalten, die zumindest durch Abschnitte der Knoten-C.R.s 1116(1)-1116(N), der gruppierten Rechenressourcen 1114 und/oder des verteilten Dateisystems 1138 der Framework-Schicht 1120 verwendet werden. Eine oder mehrere Arten von Anwendungen können eine beliebige Anzahl einer Genomikanwendung, einer kognitiven Rechenanwendung und einer maschinellen Lermanwendung umfassen, die Trainings- oder Inferenzierungssoftware beinhaltet, Framework-Software des maschinellen Lernens (z.B. PyTorch, TensorFlow, Caffe usw.) und/oder andere maschinelle Lermanwendungen beinhalten, ohne darauf beschränkt zu sein, die in Verbindung mit einer oder mehreren Ausführungsformen verwendet werden.

[0191] In zumindest einer Ausführungsform können beliebige des Konfigurationsmanagers 1134, des Ressourcenmanagers 1136 und des Ressourcen-Orchestrators 1112 eine beliebige Anzahl und Art von selbstmodifizierenden Handlungen umsetzen, die auf einer beliebigen Menge und Art von Daten basieren, die auf eine beliebige technisch machbare Weise erfasst wurden. Selbstmodifizierende Handlungen können einen Rechenzentrumsbetreiber des Rechenzentrums 1100 dahingehend entlasten, möglicherweise schlechte Konfigurationsentscheidungen zu treffen und möglicherweise nicht ausgelastete

und/oder schlecht funktionierende Abschnitte eines Rechenzentrums zu vermeiden.

[0192] Das Rechenzentrum 1100 kann Werkzeuge, Dienste, Software oder andere Ressourcen beinhalten, um ein oder mehrere Maschinenlernmodelle zu trainieren oder Informationen unter Verwendung eines oder mehrerer Maschinenlernmodelle gemäß einer oder mehreren in dieser Schrift beschriebenen Ausführungsformen vorherzusagen oder zu inferenzieren. Zum Beispiel kann/können ein Maschinenlernmodell(e) trainiert werden, indem Gewichtungssparameter gemäß einer Architektur eines neuronalen Netzwerks unter Verwendung von Software und/oder Rechenressourcen berechnet werden, die vorstehend in Bezug auf das Rechenzentrum 1100 beschrieben sind. In zumindest einer Ausführungsform können trainierte oder eingesetzte Maschinenlernmodelle, die einem oder mehreren neuronalen Netzwerken entsprechen, verwendet werden, um Informationen unter Verwendung der vorstehend in Bezug auf das Rechenzentrum 1100 beschriebenen Ressourcen zu inferenzieren oder vorherzusagen, indem Gewichtungssparameter verwendet werden, die durch eine oder mehrere Trainingstechniken berechnet werden, wie etwa die hierin beschriebenen, ohne darauf beschränkt zu sein.

[0193] In zumindest einer Ausführungsform kann das Rechenzentrum 1100 CPUs, anwendungsspezifische integrierte Schaltungen (ASICs), GPUs, FPGAs und/oder andere Hardware (oder entsprechende virtuelle Rechenressourcen) verwenden, um das Training und/oder die Inferenzierung mit den oben beschriebenen Ressourcen durchzuführen. Darüber hinaus können eine oder mehrere der vorstehend beschriebenen Software- und/oder Hardware-Ressourcen als Dienst konfiguriert sein, um Benutzern das Trainieren oder Durchführen des Inferenzierens von Informationen zu ermöglichen, wie etwa Bilderkennung, Spracherkennung oder andere Dienste der künstlichen Intelligenz.

BEISPIELHAFTE NETZWERKUMGEBUNGEN

[0194] Netzwerkumgebungen, die für die Verwendung beim Umsetzen von Ausführungsformen der Offenbarung geeignet sind, können ein oder mehrere Client-Vorrichtungen, Server, netzwerkverbundene Speicher (network attached storage - NAS), andere Backend-Vorrichtungen und/oder andere Vorrichtungstypen beinhalten. Die Client-Vorrichtungen, Server und/oder andere Vorrichtungstypen (z.B. jede Vorrichtung) können auf einer oder mehreren Instanzen der Rechenvorrichtung(en) 1000 aus **Fig. 10** umgesetzt sein - z.B. kann jede Vorrichtung ähnliche Komponenten, Merkmale und/oder Funktionalität der Rechenvorrichtung(en) 1000 beinhalten. Zusätzlich, wenn Backend-Vorrichtungen (z.B. Server, NAS usw.) umgesetzt werden, können die

Backend-Vorrichtungen auch Teil eines Rechenzentrums 1100 sein, dessen Beispiel in **Fig. 11** näher beschrieben wird.

[0195] Komponenten einer Netzwerkumgebung können miteinander über ein oder mehrere Netzwerke kommunizieren, die drahtgebunden, drahtlos oder beides sein können. Das Netzwerk kann mehrere Netzwerke oder ein Netzwerk von Netzwerken beinhalten. Zum Beispiel kann das Netzwerk ein oder mehrere Weitverkehrsnetzwerke (WAN), ein oder mehrere lokale Netzwerke (LANs), ein oder mehrere öffentliche Netzwerke wie das Internet und/oder ein öffentliches Telefonvermittlungsnetz (public switched telephone network - PSTN) und/oder ein oder mehrere private Netzwerke beinhalten. Wenn das Netzwerk ein drahtloses Telekommunikationsnetz beinhaltet, können Komponenten wie eine Basisstation, ein Kommunikationsturm oder sogar Zugangspunkte (sowie andere Komponenten) eine drahtlose Konnektivität bereitstellen.

[0196] Kompatible Netzwerkumgebungen können eine oder mehrere Peer-to-Peer-Netzwerkumgebungen beinhalten - in diesem Fall kann ein Server nicht in einer Netzwerkumgebung beinhaltet sein - und eine oder mehrere Client-Server-Netzwerkumgebungen - in diesem Fall können ein oder mehrere Server in einer Netzwerkumgebung beinhaltet sein. In Peer-to-Peer-Netzwerkumgebungen kann die hierin in Bezug auf einen oder mehrere Server beschriebene Funktionalität auf einer beliebigen Anzahl von Client-Vorrichtungen umgesetzt sein.

[0197] In zumindest einer Ausführungsform kann eine Netzwerkumgebung eine oder mehrere Cloud-basierte Netzwerkumgebungen, eine verteilte Rechenumgebung, eine Kombination davon usw. beinhalten. Eine Cloud-basierte Netzwerkumgebung kann eine Framework-Schicht, einen Job-Scheduler, einen Ressourcenmanager und ein verteiltes Dateisystem beinhalten, die auf einem oder mehreren Servern umgesetzt sind, die einen oder mehrere Kernnetzwerkserver und/oder Edge-Server beinhalten können. Eine Framework-Schicht kann ein Framework zur Unterstützung von Software einer Software-Schicht und/oder einer oder mehrerer Anwendungen einer Anwendungsschicht beinhalten. Die Software oder Anwendung(en) können jeweils Web-basierte Dienstsoftware oder Anwendungen beinhalten. In Ausführungsformen können eine oder mehrere der Client-Vorrichtungen die Web-basierte Dienstsoftware oder Anwendungen verwenden (z.B. durch Zugreifen auf die Dienstsoftware und/oder Anwendungen über eine oder mehrere Anwendungsprogrammierschnittstellen (API)). Bei der Framework-Schicht kann es sich um eine Art freies und quelloffenes Software-Webanwendungs-Framework handeln, das etwa ein verteiltes Dateisystem für die

Verarbeitung großer Datenmengen (z.B. „Big Data“) verwenden kann, ist aber nicht darauf beschränkt.

[0198] Eine Cloud-basierte Netzwerkkumgebung kann Cloud-Computing und/oder Cloud-Speicher bereitstellen, die eine beliebige Kombination von hierin beschriebenen Rechen- und/oder Datenspeicherfunktionen (oder einen oder mehrere Teile davon) ausführen. Jede dieser verschiedenen Funktionen kann über mehrere Standorte von zentralen oder Kernservern (z.B. von einem oder mehreren Rechenzentren, die über einen Staat, eine Region, ein Land, den Globus usw. verteilt sein können) verteilt sein. Wenn eine Verbindung zu einem Benutzer (z.B. einer Client-Vorrichtung) relativ nahe bei einem oder mehreren Edge-Servern ist, können ein oder mehrere Core-Server dem oder den Edge-Servern zumindest einen Teil der Funktionalität zuweisen. Eine Cloud-basierte Netzwerkkumgebung kann privat sein (z.B. auf eine einzelne Organisation beschränkt), kann öffentlich sein (z.B. für viele Organisationen verfügbar) und/oder eine Kombination davon sein (z.B. eine hybride Cloud-Umgebung).

[0199] Die Client-Vorrichtung(en) können zumindest eine der Komponenten, Merkmale und Funktionalität der hierin in Bezug auf **Fig. 10** beschriebenen Rechen- vorrichtung(en) 1000 beinhalten. Als Beispiel und nicht einschränkend kann eine Client-Vorrichtung als Personal Computer (PC), Laptop-Computer, mobile Vorrichtung, Smartphone, Tablet-Computer, Smartwatch, tragbarer Computer, Personal Digital Assistant (PDA), MP3-Player, Virtual-Reality-Headset, System oder Vorrichtung zur globalen Positionsbestimmung (GPS), Videoplayer, Videokamera, Überwachungsvorrichtung oder -system, Fahrzeug, Boot, Flugschiff, virtuelle Maschine, Drohne, Roboter, tragbare Kommunikationsvorrichtung, Vorrichtung in einem Krankenhaus, Spielgerät oder -system, Unterhaltungssystem, Fahrzeugcomputersystem, eingebetteter Systemcontroller, Fernbedienung, Haushaltsgerät, Unterhaltungselektronikgerät, Workstation, Edge-Vorrichtung, eine beliebige Kombination dieser skizzierten Vorrichtungen oder jede andere geeignete Vorrichtung verkörpert sein.

[0200] Die Offenbarung kann im allgemeinen Kontext von Computercode- oder maschinenverwendbaren Anweisungen, einschließlich computerausführbarer Anweisungen wie Programmmodulen, die von einem Computer oder einem anderen Computer, wie einem Personal Data Assistent oder einem anderen Handheld-Gerät, ausgeführt werden, beschrieben werden. Im Allgemeinen beziehen sich Programmmodule einschließlich Routinen, Programme, Objekte, Komponenten, Datenstrukturen usw. auf Code, der bestimmte Aufgaben ausführt oder bestimmte abstrakte Datentypen umgesetzt. Die Offenbarung kann in einer Vielzahl von Systemkonfi-

gurationen praktiziert werden, einschließlich Handheld-Vorrichtungen, Unterhaltungselektronik, Allzweckcomputern, speziellen Rechen- vorrichtungen usw. Die Offenbarung kann auch in verteilten Rechenumgebungen praktiziert werden, in denen Aufgaben von entfernten Verarbeitungsvorrichtungen, die über ein Kommunikationsnetzwerk verbunden sind, durchgeführt werden.

[0201] Wie hierin verwendet, sollte eine Rezipitation von „und/oder“ in Bezug auf zwei oder mehr Elemente so ausgelegt werden, dass sie nur ein Element oder eine Kombination von Elementen bedeutet. Zum Beispiel kann „Element A, Element B und/oder Element C“ nur Element A, nur Element B, nur Element C, Element A und Element B, Element A und Element C, Element B und Element C oder Elemente enthalten A, B und C. Außerdem kann „zumindest eines von Element A oder Element B“ zumindest eines von Element A, zumindest eines von Element B oder zumindest eines von Element A und zumindest eines von Element B umfassen. Ferner kann „zumindest eines von Element A und Element B“ zumindest eines von Element A, zumindest eines von Element B oder zumindest eines von Element A und zumindest eines von Element B beinhalten.

[0202] Der Gegenstand der vorliegenden Offenbarung wird hierin spezifisch beschrieben, um gesetzliche Anforderungen zu erfüllen. Die Beschreibung selbst soll jedoch den Umfang dieser Offenbarung nicht einschränken. Vielmehr haben die Erfinder in Erwägung gezogen, dass der beanspruchte Gegenstand auch auf andere Weise verkörpert werden könnte, um andere Schritte oder Kombinationen von Schritten ähnlich den in diesem Dokument beschriebenen in Verbindung mit anderen gegenwärtigen oder zukünftigen Technologien einzuschließen. Obwohl die Begriffe „Schritt“ und/oder „Block“ hierin verwendet werden können, um verschiedene Elemente der verwendeten Verfahren zu bezeichnen, sollten die Begriffe darüber hinaus nicht so ausgelegt werden, dass sie eine bestimmte Reihenfolge zwischen oder zwischen verschiedenen hierin offenbarten Schritten implizieren, es sei denn, die Reihenfolge ist der einzelnen Schritte ist explizit beschrieben.

ZITATE ENTHALTEN IN DER BESCHREIBUNG

Diese Liste der vom Anmelder aufgeführten Dokumente wurde automatisiert erzeugt und ist ausschließlich zur besseren Information des Lesers aufgenommen. Die Liste ist nicht Bestandteil der deutschen Patent- bzw. Gebrauchsmusteranmeldung. Das DPMA übernimmt keinerlei Haftung für etwaige Fehler oder Auslassungen.

Zitierte Patentliteratur

- US 16101232 [0105]

Patentansprüche**1. Verfahren, umfassend:**

Erlangen von Textdaten, die eine erste Frage darstellen, die einer Maschine zugeordnet ist;
Bestimmen, zumindest auf Grundlage der Textdaten, zumindest einer von einer oder mehreren zweiten Fragen, die der ersten Frage zugeordnet sind, oder einer oder mehreren Antworten, die der ersten Frage zugeordnet sind;
Bestimmen, unter Verwendung eines oder mehrerer Sprachmodelle und zumindest auf Grundlage der Textdaten und der Daten, die eine oder mehrere zweite Fragen und/oder eine oder mehrere Antworten darstellen, einer Ausgabe, die der ersten Frage zugeordnet ist; und
Bewirken, dass die Ausgabe unter Verwendung einer oder mehrerer Komponenten der Maschine kommuniziert wird.

2. Verfahren nach Anspruch 1, wobei das eine oder die mehreren Sprachmodelle ein großes Sprachmodell (LLM) und/oder ein generatives Sprachmodell und/oder ein generatives vortrainiertes Transformationsmodell und/oder ein generatives Transformationsmodell aufweisen.

3. Verfahren nach Anspruch 1 oder 2, ferner umfassend:

Bestimmen von Kontextinformationen, die der ersten Frage zugeordnet sind, zumindest auf Grundlage der Textdaten,
wobei das Bestimmen der Ausgabe, die der ersten Frage zugeordnet ist, ferner zumindest auf Daten basiert, welche die Kontextinformationen darstellen.

4. Verfahren nach Anspruch 3, wobei das Bestimmen der der ersten Frage zugeordneten Kontextinformationen das Bestimmen, zumindest auf Grundlage der Textdaten umfasst, dass zumindest ein Teil einer der Maschine entsprechenden Wissensbasis der ersten Frage zugeordnet ist, wobei zumindest der Teil der Wissensbasis den Kontextinformationen zugeordnet ist.

5. Verfahren nach Anspruch 4, wobei das eine oder die mehreren Sprachmodelle ein oder mehrere feste Sprachmodelle aufweisen und die Wissensbasis eine Live-Wissensbasis aufweist.

6. Verfahren nach einem der vorstehenden Ansprüche, wobei das Bestimmen der einen oder mehreren zweiten Fragen, die der ersten Frage zugeordnet sind, oder der einen oder mehreren Antworten, die der ersten Frage zugeordnet sind, umfasst:

Bestimmen, zumindest auf Grundlage der Textdaten, eines oder mehrerer Frage-Antwort-Paare, die sich auf die erste Frage beziehen, wobei zumindest ein einzelnes Frage-Antwort-Paar des einen oder

der mehreren Frage-Antwort-Paare eine zweite Frage aus der einen oder den mehreren zweiten Fragen und eine entsprechende Antwort aus der einen oder den mehreren Antworten enthält.

7. Verfahren nach einem der vorstehenden Ansprüche, wobei das Bestimmen der einen oder mehreren zweiten Fragen, die der ersten Frage zugeordnet sind, und/oder der einen oder mehreren Antworten, die der ersten Frage zugeordnet sind, umfasst:

Erzeugen einer ersten Einbettung, die der ersten Frage zugeordnet ist, zumindest auf Grundlage der Textdaten;

Analysieren der ersten Einbettung in Bezug auf eine oder mehrere zweite Einbettungen, die der einen oder mehreren zweiten Fragen und/oder der einen oder mehreren Antworten zugeordnet sind;

Bestimmen, zumindest auf Grundlage des Analysierens, dass zumindest eine zweite Einbettung der einen oder mehreren zweiten Einbettungen der ersten Einbettung ähnlich ist; und

Bestimmen, dass die zweite Einbettung einer zweiten Frage der einen oder mehreren zweiten Fragen und/oder einer Antwort der einen oder mehreren Antworten zugeordnet ist.

8. Verfahren nach einem der vorstehenden Ansprüche, ferner umfassend:

Erzeugen von Eingabeaufforderungsdaten, die eine Eingabeaufforderung darstellen, zumindest auf Grundlage der Textdaten und der Daten, die eine oder mehrere zweite Fragen und/oder eine oder mehrere Antworten darstellen,
wobei das Bestimmen der der ersten Frage zugeordneten Ausgabe zumindest darauf basiert, dass die Eingabeaufforderungsdaten unter Verwendung des einen oder der mehreren Sprachmodelle verarbeitet werden.

9. Verfahren nach einem der Ansprüche 6 bis 8, wobei:

ein erster Teil der Eingabeaufforderung die eine oder mehrere zweite Fragen und/oder die eine oder mehrere Antworten umfasst; und

ein zweiter Teil der Eingabeaufforderung die erste Frage umfasst, wobei der zweite Teil in der Eingabeaufforderung hinter dem ersten Teil liegt.

10. Verfahren nach einem der vorstehenden Ansprüche, ferner umfassend:

vor dem Bestimmen der Ausgabe, Bestimmen einer zweiten Ausgabe, die der dritten Frage zugeordnet ist, unter Verwendung des einen oder der mehreren Sprachmodelle und zumindest auf Grundlage von zweiten Textdaten, die eine dritte, der Maschine zugeordnete Frage darstellen,

wobei das Bestimmen der Ausgabe, die der ersten Frage zugeordnet ist, ferner auf zumindest dem

einen oder den mehreren Sprachmodellen basiert, welche die zweite Ausgabe verarbeiten.

11. Verfahren nach einem der vorstehenden Ansprüche, wobei:
die erste Frage einer Komponente der Maschine und/oder einem Merkmal der Maschine und/oder einer der Maschine zugeordneten Wartung zugeordnet ist; und
die Ausgabe Informationen darstellt, die der Komponente der Maschine und/oder dem Merkmal der Maschine und/oder der der Maschine zugeordneten Wartung zugeordnet sind.

12. System, aufweisend:
eine oder mehrere Verarbeitungseinheiten, um:
Textdaten zu erzeugen, die eine einer Maschine zugeordnete Frage darstellen;
zumindest auf Grundlage der Textdaten und unter Verwendung einer der Maschine entsprechenden Wissensbasis die der Frage zugeordneten Kontextinformationen zu bestimmen;
unter Verwendung eines oder mehrerer Sprachmodelle und zumindest auf Grundlage der Textdaten und der Daten, welche die Kontextinformationen darstellen, eine der Frage zugeordnete Ausgabe zu bestimmen; und
Kommunikation der Ausgabe unter Verwendung einer oder mehrerer Komponenten der Maschine zu bewirken.

13. System nach Anspruch 12, wobei die der Maschine entsprechende Wissensbasis eines oder mehrere der folgenden Elemente aufweist:
Informationen aus einem der Maschine zugeordneten Benutzerhandbuch; oder
Informationen aus einem oder mehreren Benutzerhandbüchern, die einer oder mehreren anderen Maschinen zugeordnet sind.

14. System nach Anspruch 12 oder 13, wobei die eine oder die mehreren Verarbeitungseinheiten ferner dazu dienen, um:
zumindest auf Grundlage der Textdaten eine oder mehrere Fragen, die der Frage zugeordnet sind, und eine oder mehrere Antworten, die der einen oder den mehreren Fragen zugeordnet sind, zu bestimmen,
wobei die Bestimmung der Ausgabe, die der Frage zugeordnet ist, ferner zumindest auf Daten basiert, welche die eine oder mehreren Fragen und die eine oder mehreren Antworten darstellen.

15. System nach einem der Ansprüche 12 bis 14, wobei die der Frage zugeordnete Kontextinformation bestimmt wird zumindest durch:
Erzeugen einer ersten Einbettung, die der Frage zugeordnet ist, zumindest auf Grundlage der Textdaten;
Analysieren der ersten Einbettung in Bezug auf eine

oder mehrere zweite Einbettungen, die einem oder mehreren Teilen der Wissensbasis zugeordnet sind;
Bestimmen, zumindest auf Grundlage des Analysierens, dass zumindest eine zweite Einbettung der einen oder mehreren zweiten Einbettungen der ersten Einbettung ähnlich ist; und
Bestimmen, dass die zweite Einbettung einem Teil der Wissensbasis zugeordnet ist, wobei der Teil der Wissensbasis den kontextbezogenen Informationen entspricht.

16. System nach einem der Ansprüche 12 bis 15, wobei die eine oder die mehreren Verarbeitungseinheiten ferner dazu dienen, um:
zumindest auf Grundlage der Textdaten und der Daten, welche die Kontextinformationen darstellen, Eingabeaufforderungsdaten zu erzeugen, die eine Eingabeaufforderung darstellen,
wobei die der Frage zugeordnete Ausgabe zumindest auf Grundlage der Eingabeaufforderungsdaten bestimmt wird.

17. System nach Anspruch 16, wobei:
ein erster Teil der Eingabeaufforderung die Kontextinformationen umfasst; und
ein zweiter Teil der Eingabeaufforderung die Frage umfasst, wobei der zweite Teil in der Eingabeaufforderung hinter dem ersten Teil liegt.

18. System nach einem der Ansprüche 12 bis 17, wobei die eine oder die mehreren Verarbeitungseinheiten ferner dazu dienen, um:
vor der Bestimmung der Ausgabe unter Verwendung des einen oder der mehreren Sprachmodelle und zumindest auf Grundlage von zweiten Textdaten, die eine zweite, der Maschine zugeordnete Frage darstellen, eine zweite, der zweiten Frage zugeordnete Ausgabe zu bestimmen,
wobei die der Frage zugeordnete Ausgabe zumindest auf Grundlage der zweiten Ausgabe bestimmt wird.

19. System nach einem der Ansprüche 12 bis 18, wobei:
die Frage einer Komponente der Maschine und/oder einem Merkmal der Maschine und/oder einer der Maschine zugeordneten Wartung zugeordnet ist; und
die Ausgabe eine Antwort darstellt, die Informationen enthält, die einer Komponente der Maschine und/oder dem Merkmal der Maschine und/oder der der Maschine zugeordneten Wartung zugeordnet sind.

20. System nach einem der Ansprüche 12 bis 19, wobei das System enthalten ist in:
einem Steuerungssystem für eine autonome oder halbautonome Maschine; und/oder
einem Wahrnehmungssystem für eine autonome oder halbautonome Maschine; und/oder

einem System zur Durchführung von Simulationsoptionen; und/oder
 einem System zur Durchführung digitaler Twinning-Operationen; und/oder
 einem System zur Durchführung von Echtzeit-Streaming; und/oder
 einem System zur Erzeugung oder Darstellung von Inhalten der virtuellen Realität (VR); und/oder
 einem System zur Erzeugung oder Darstellung von Inhalten der erweiterten Realität (AR); und/oder
 einem System zur Erzeugung oder Darstellung von Inhalten der gemischten Realität (MR); und/oder
 einem System zur Durchführung von Lichttransport-simulationen; und/oder
 einem System zur kollaborativen Erstellung von Inhalten für 3D-Assets; und/oder
 einem System zur Durchführung von Deep-Learning-Operationen; und/oder
 einem System, das vermittelt einer Edge-Vorrichtung implementiert ist; und/oder
 einem System, das vermittelt eines Roboters implementiert ist; und/oder
 einem System zur Durchführung von KI-Gesprächs-Operationen; und/oder
 einem System zur Erzeugung synthetischer Daten; und/oder
 einem System, das eine oder mehrere virtuelle Maschinen (VMs) aufweist; und/oder
 einem System, das zumindest teilweise in einem Rechenzentrum implementiert ist; und/oder
 einem System, das zumindest teilweise vermittelt Cloud-Computing-Ressourcen implementiert ist.

Inhalten der virtuellen Realität (VR); und/oder
 einem System zur Erzeugung oder Darstellung von Inhalten der erweiterten Realität (AR); und/oder
 einem System zur Erzeugung oder Darstellung von Inhalten der gemischten Realität (MR); und/oder
 einem System zur Durchführung von Lichttransport-simulationen; und/oder
 einem System zur kollaborativen Erstellung von Inhalten für 3D-Assets; und/oder
 einem System zur Durchführung von Deep-Learning-Operationen; und/oder
 einem System, das vermittelt einer Edge-Vorrichtung implementiert ist; und/oder
 einem System, das vermittelt eines Roboters implementiert ist; und/oder
 einem System zur Durchführung von KI-Gesprächs-Operationen; und/oder
 einem System zur Erzeugung synthetischer Daten; und/oder
 einem System, das eine oder mehrere virtuelle Maschinen (VMs) aufweist; und/oder
 einem System, das zumindest teilweise in einem Rechenzentrum implementiert ist; und/oder
 einem System, das zumindest teilweise vermittelt Cloud-Computing-Ressourcen implementiert ist.

Es folgen 14 Seiten Zeichnungen

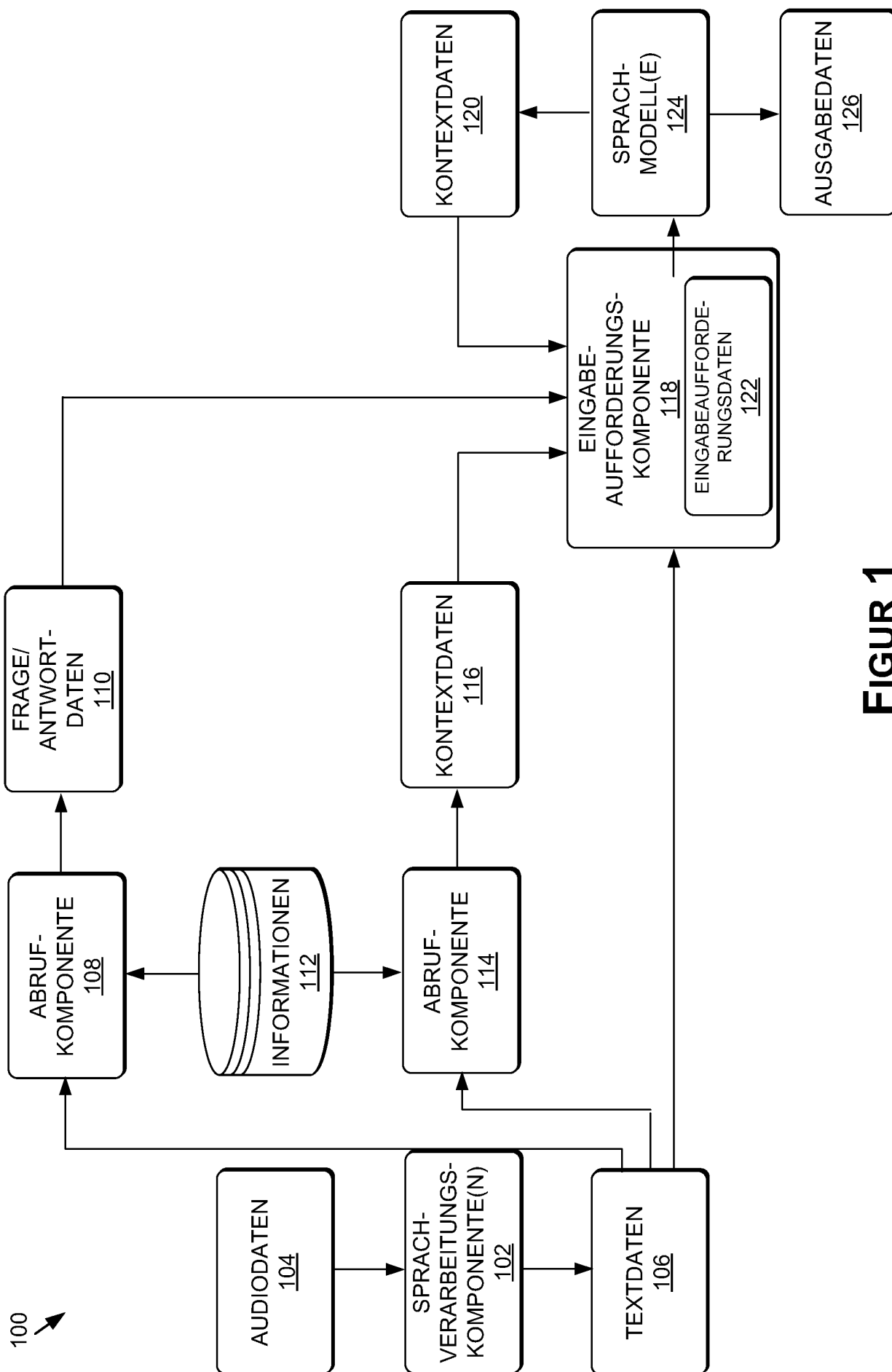
21. Prozessor, aufweisend:

eine oder mehrere Verarbeitungseinheiten, um eine Antwort auf eine Frage unter Verwendung einer oder mehrerer Komponenten einer Maschine zu übermitteln, wobei die Antwort auf Grundlage von zumindest einem oder mehreren Sprachmodellen bestimmt wird, die Textdaten, die eine einer Maschine zugeordnete Frage darstellen, Daten, die ein der Maschine zugeordnetes Frage-Antwort-Paar darstellen, sowie Daten verarbeiten, die Kontextinformationen darstellen, die unter Verwendung einer der Maschine zugeordneten Wissensbasis bestimmt werden.

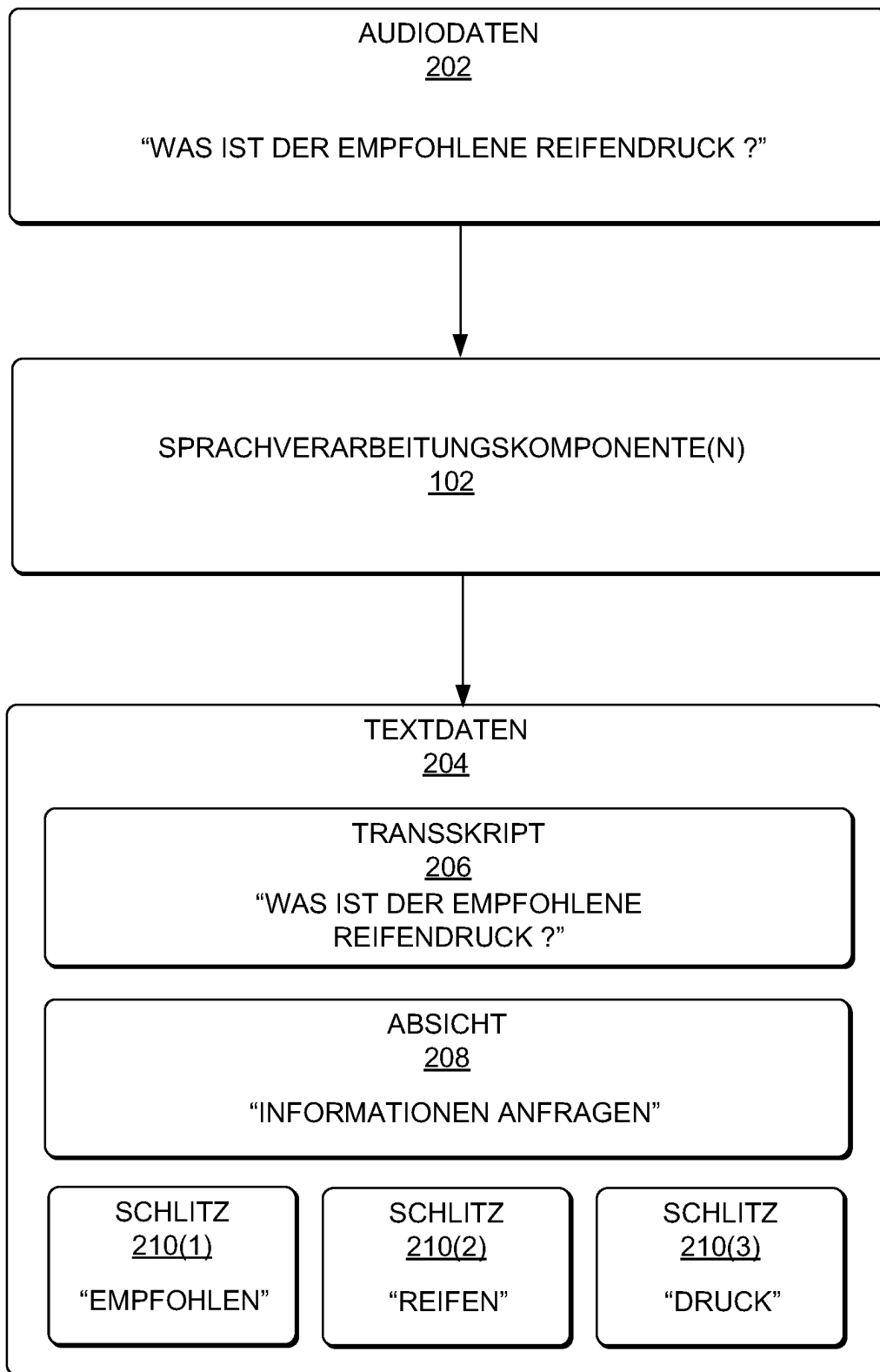
22. Prozessor nach Anspruch 21, wobei der Prozessor enthalten ist in:

einem Steuerungssystem für eine autonome oder halbautonome Maschine; und/oder
 einem Wahrnehmungssystem für eine autonome oder halbautonome Maschine; und/oder
 einem System zur Durchführung von Simulationsoptionen; und/oder
 einem System zur Durchführung digitaler Twinning-Operationen; und/oder
 einem System zur Durchführung von Echtzeit-Streaming; und/oder
 einem System zur Erzeugung oder Darstellung von

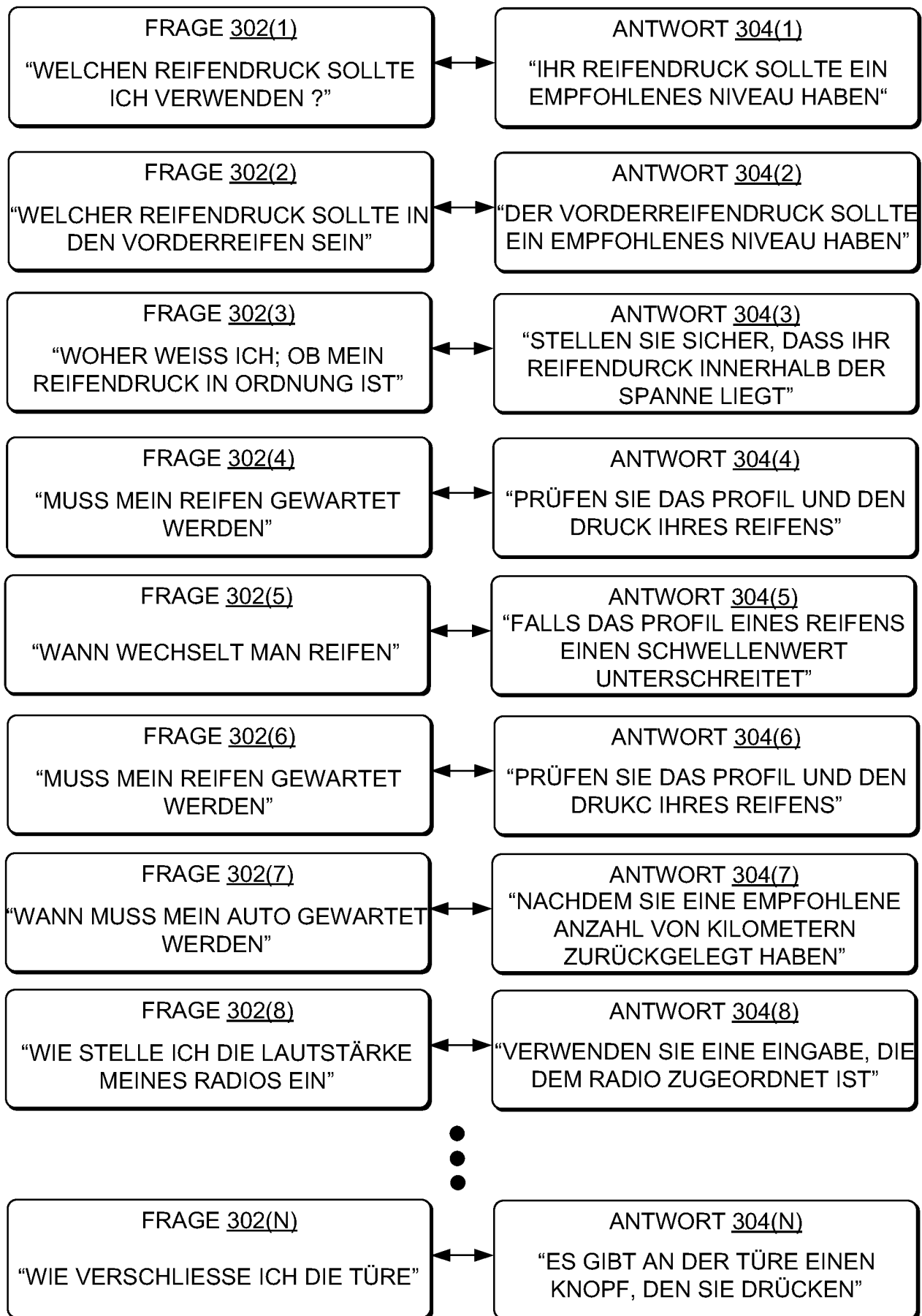
Anhängende Zeichnungen



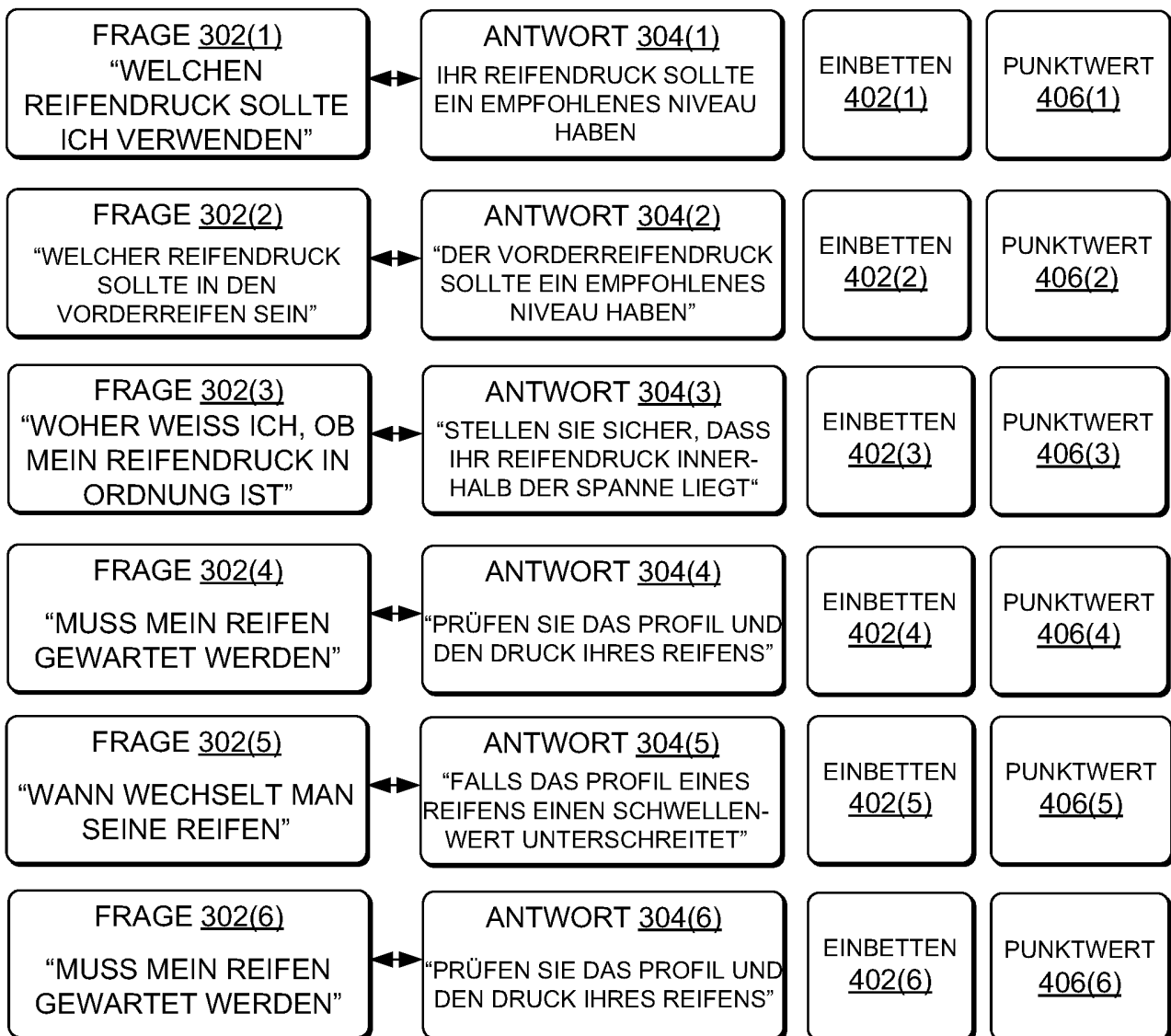
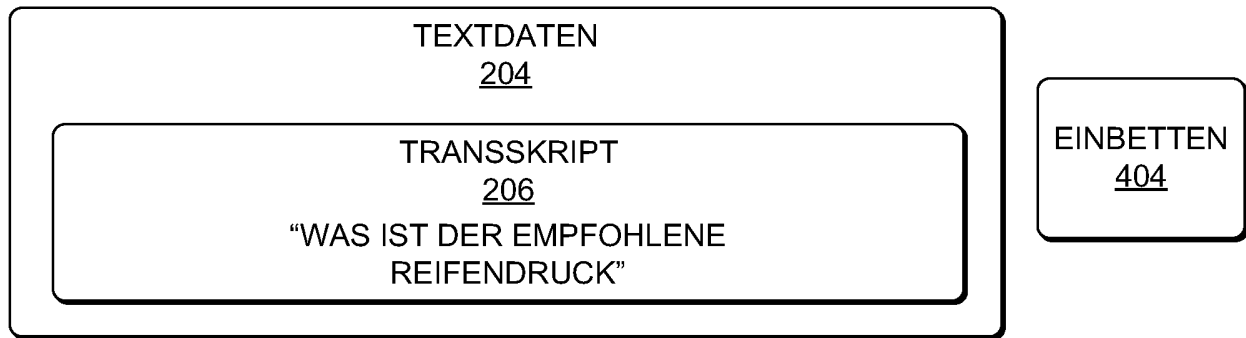
FIGUR 1



FIGUR 2



FIGUR 3



FIGUR 4

INFORMATIONEN

504 ↘

HANDBUCH

502

“PRÜFEN SIE IHRE REIFEN REGELMÄSSIG, UM SICHERZUSTELLEN, DASS SIE REIFEN ZUM FAHREN GEEIGNET SIND. ES WIRD EMPFOHLEN, DASS DIE REIFEN ZUMNINDEST JEDE WOCHEN GEPRÜFT WERDEN. BEIM PRÜFEN DER REIFEN STELLEN SIE SICHER, DASS DER REIFENDRUCK DAS EMPFOHELENE NIVEAU HAT, WELCHER BEI 2,5 BAR LIEGT. ES WIRD EBENFALLS EMPFOHLEN, DASS DIE REIFEN ZUMINDEST EINMAL IM JAHR VON EINEM FACHMANN GEPRÜFT WERDEN.”

ABSCHNITT 506(1)

“PRÜFEN SIE IHRE REIFEN REGELMÄSSIG, UM SICHERZUSTELLEN, DASS DIE REIFEN”

EINBETTEN
508(1)

PUNKTZAHL
510(1)

ABSCHNITT 506(2)

“STELLEN SIE SICHER, DASS DIE REIFEN ZUM FAHREN GEEIGNET SIND”

EINBETTEN
508(2)

PUNKTZAHL
510(2)

ABSCHNITT 506(3)

“GEEIGNET ZUM FAHREN WIRD EMPFOHLEN, DASS DIE REIFEN”

EINBETTEN
508(3)

PUNKTZAHL
510(3)

ABSCHNITT 506(4)

“EMPFOHLEN WIRD, DASS DIE REIFEN ZUMINDEST WÖCHENTLICH GEPRÜFT WERDEN”

EINBETTEN
508(4)

PUNKTZAHL
510(4)

ABSCHNITT 506(5)

“GEPRÜFT ZUMINDEST JEDE WOCHEN BEIM PRÜFEN IHRER REIFEN”

EINBETTEN
508(5)

PUNKTZAHL
510(5)

ABSCHNITT 506(6)

“BEIM PRÜFEN IHRER REIFEN STELLEN SIE SICHER, DASS DER REIFENDRUCK”

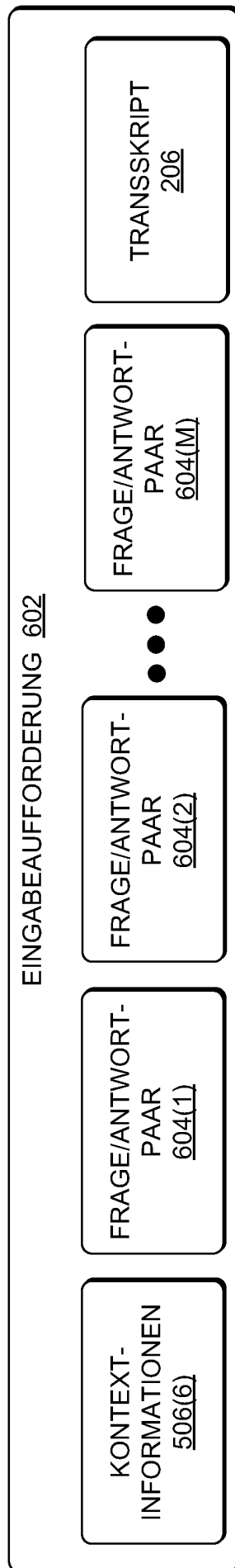
EINBETTEN
508(6)

PUNKTZAHL
510(6)



FIGUR 5

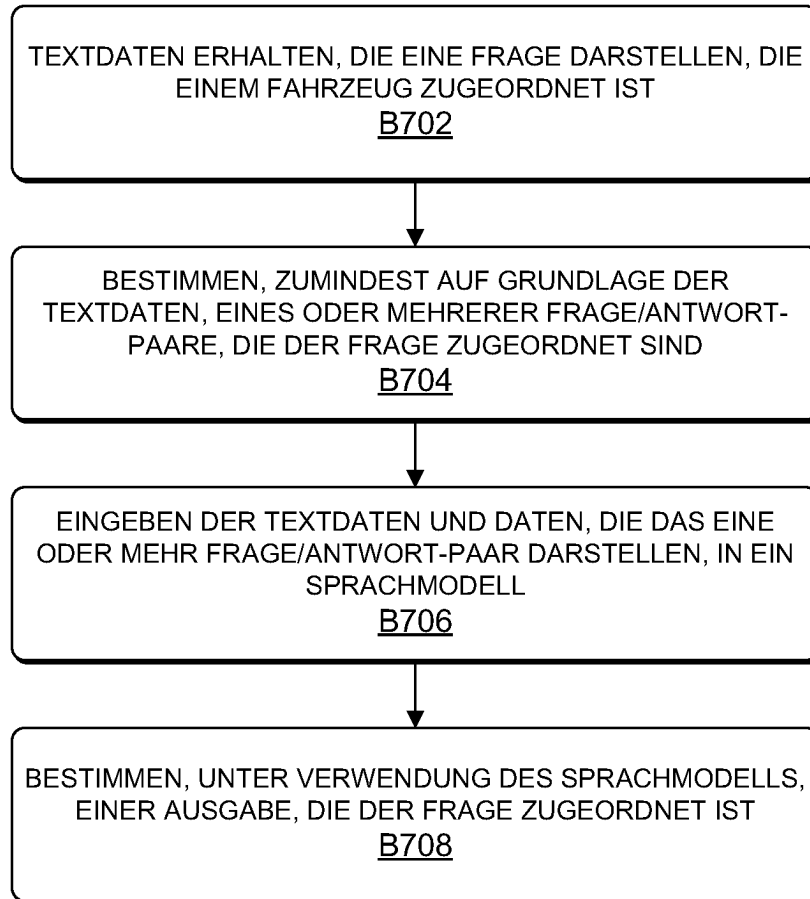




FIGUR 6

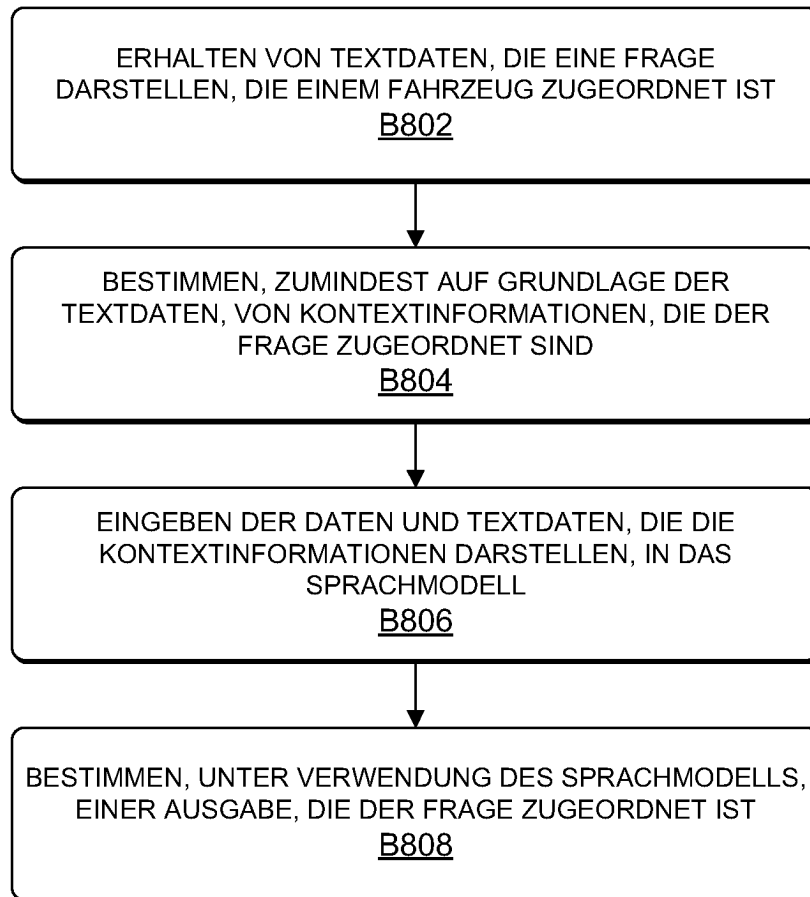
+

700
↓

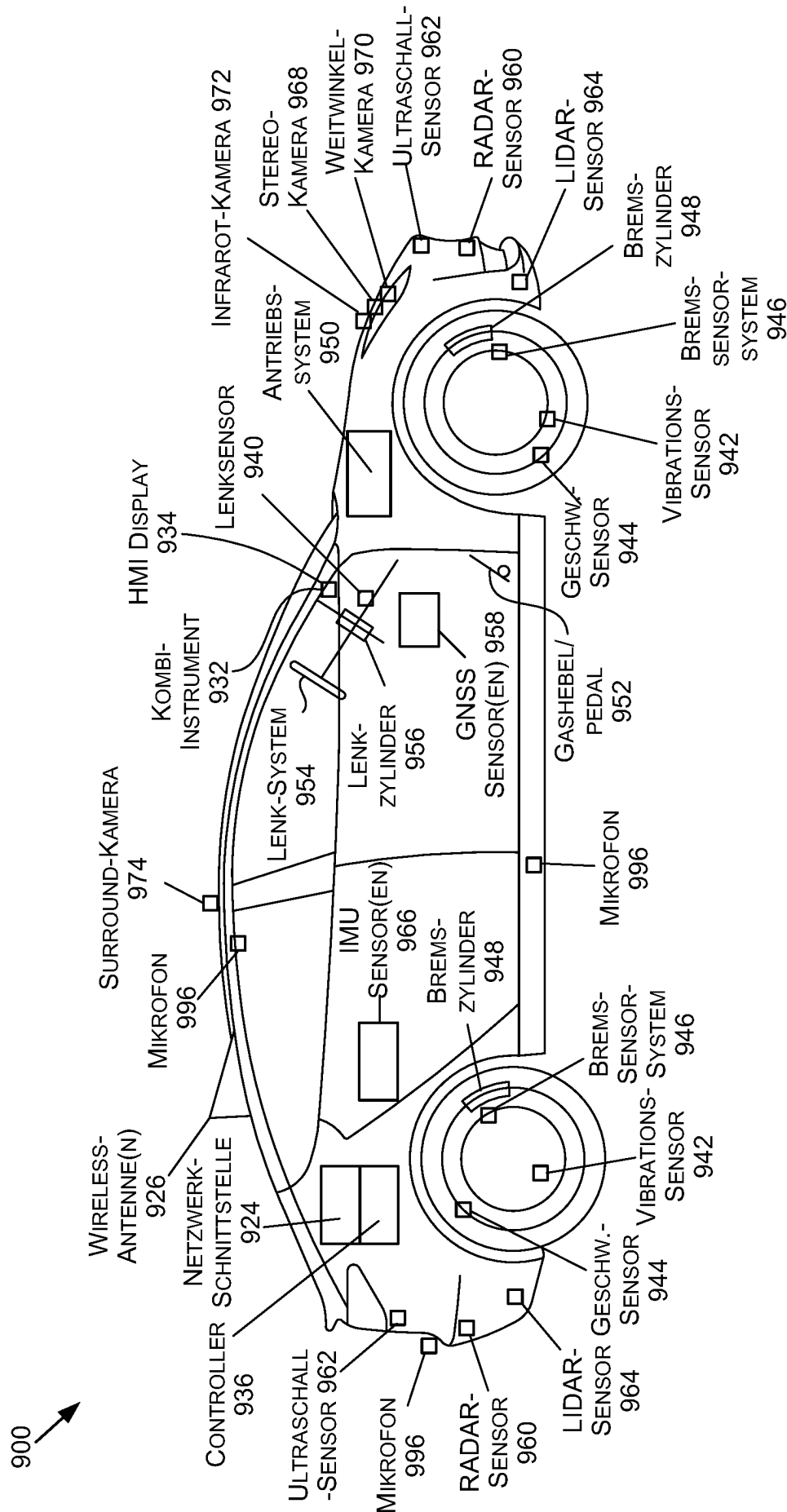


FIGUR 7

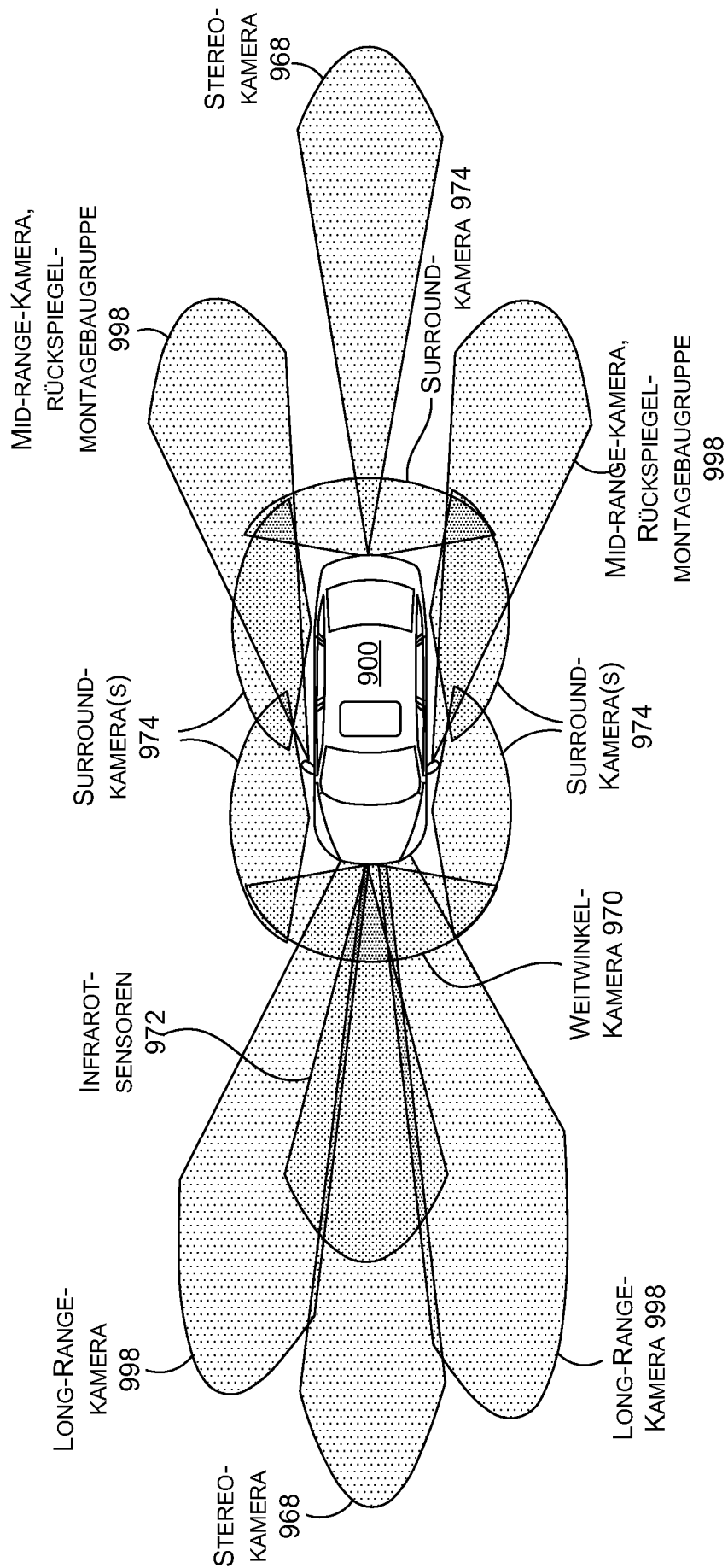
800
↓



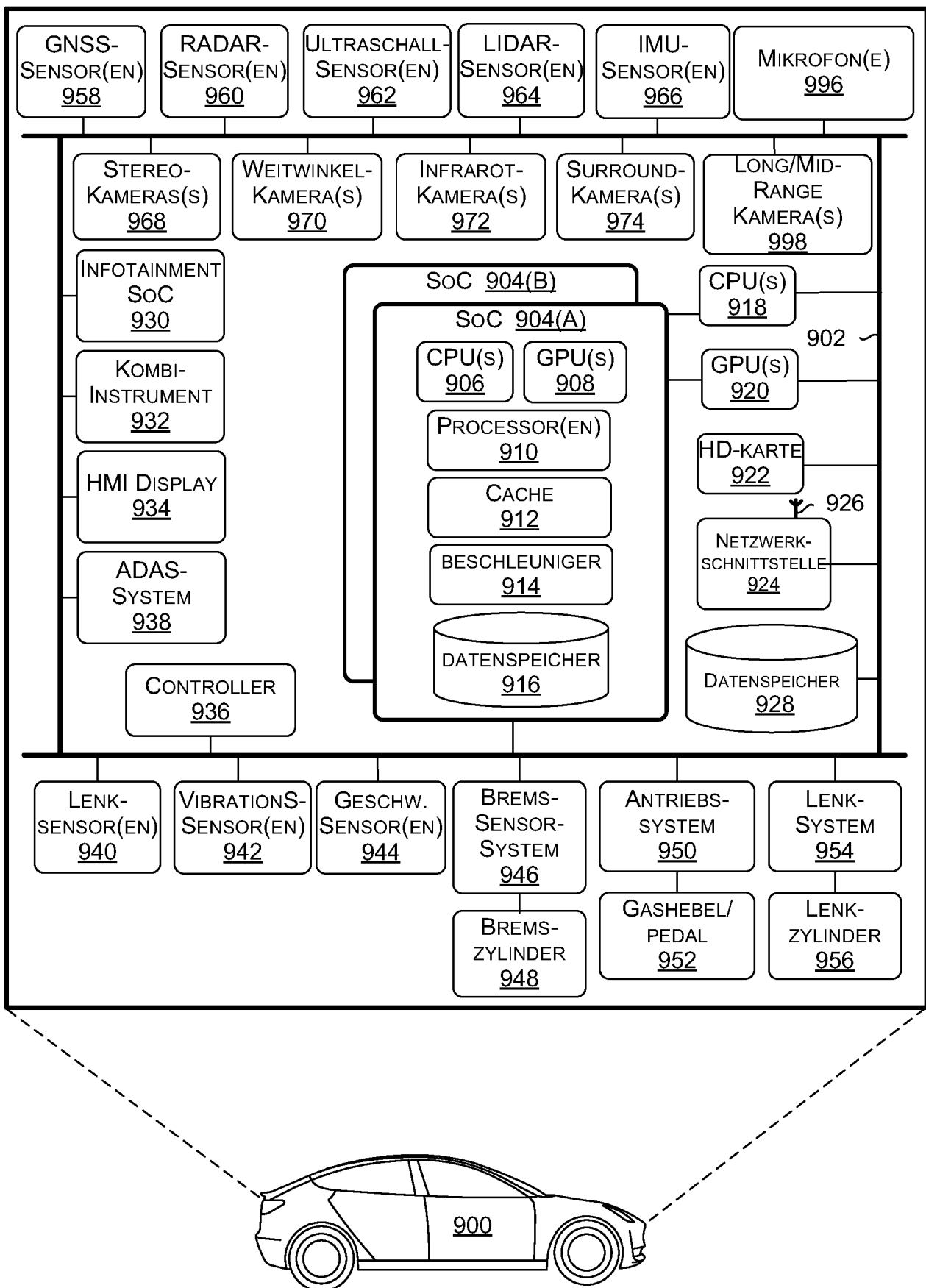
FIGUR 8



FIGUR 9A

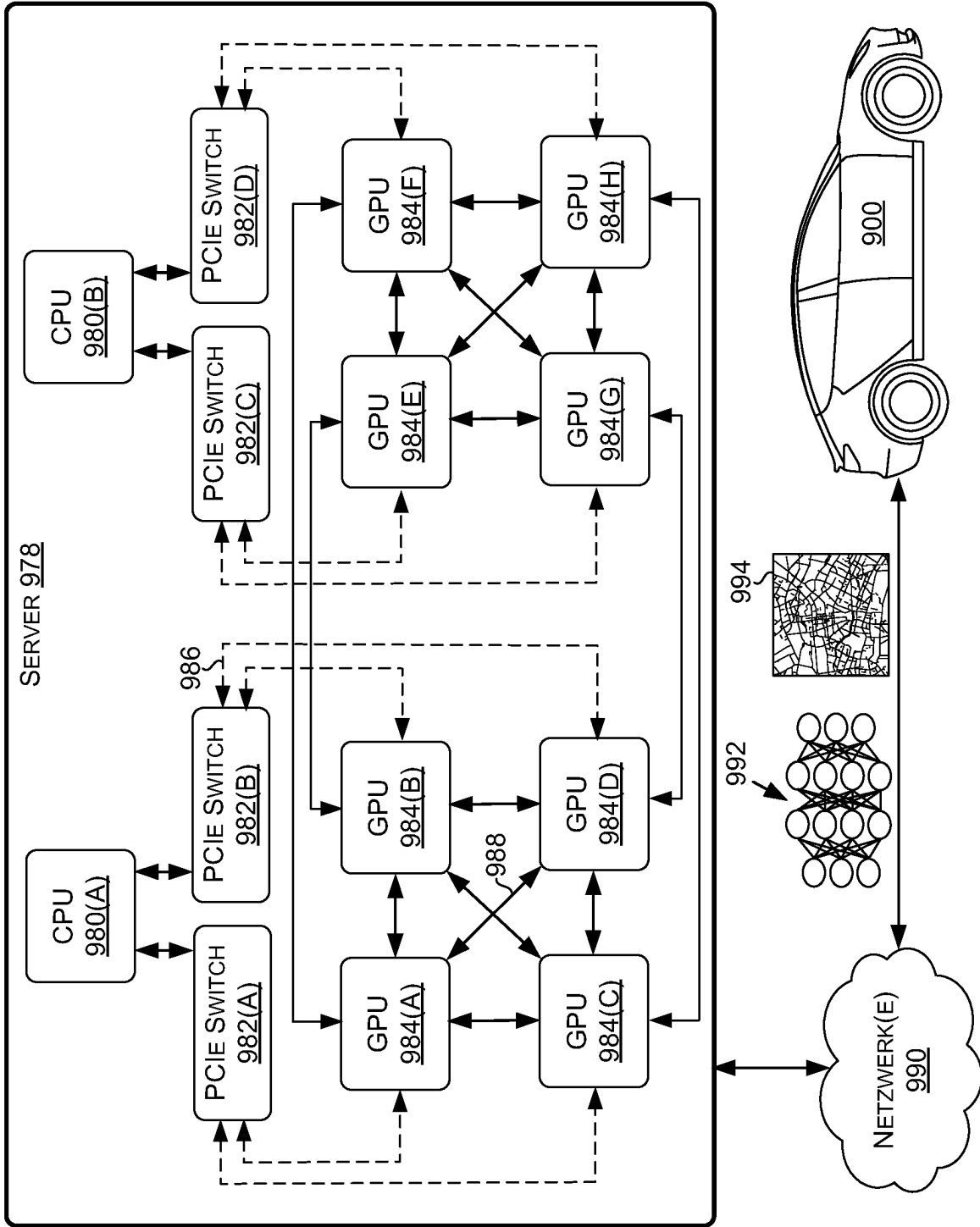


FIGUR 9B

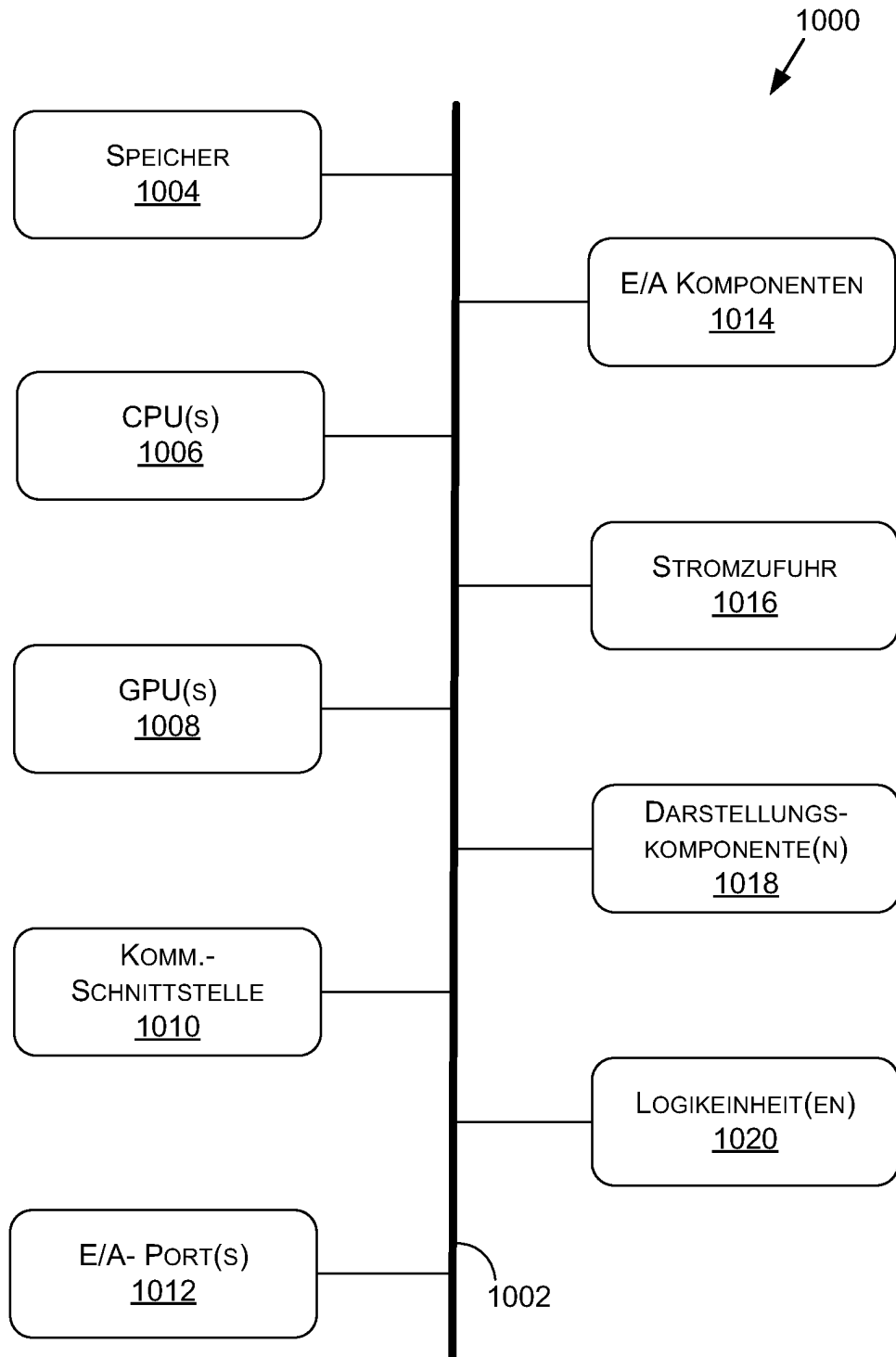


FIGUR 9C

976 ↗

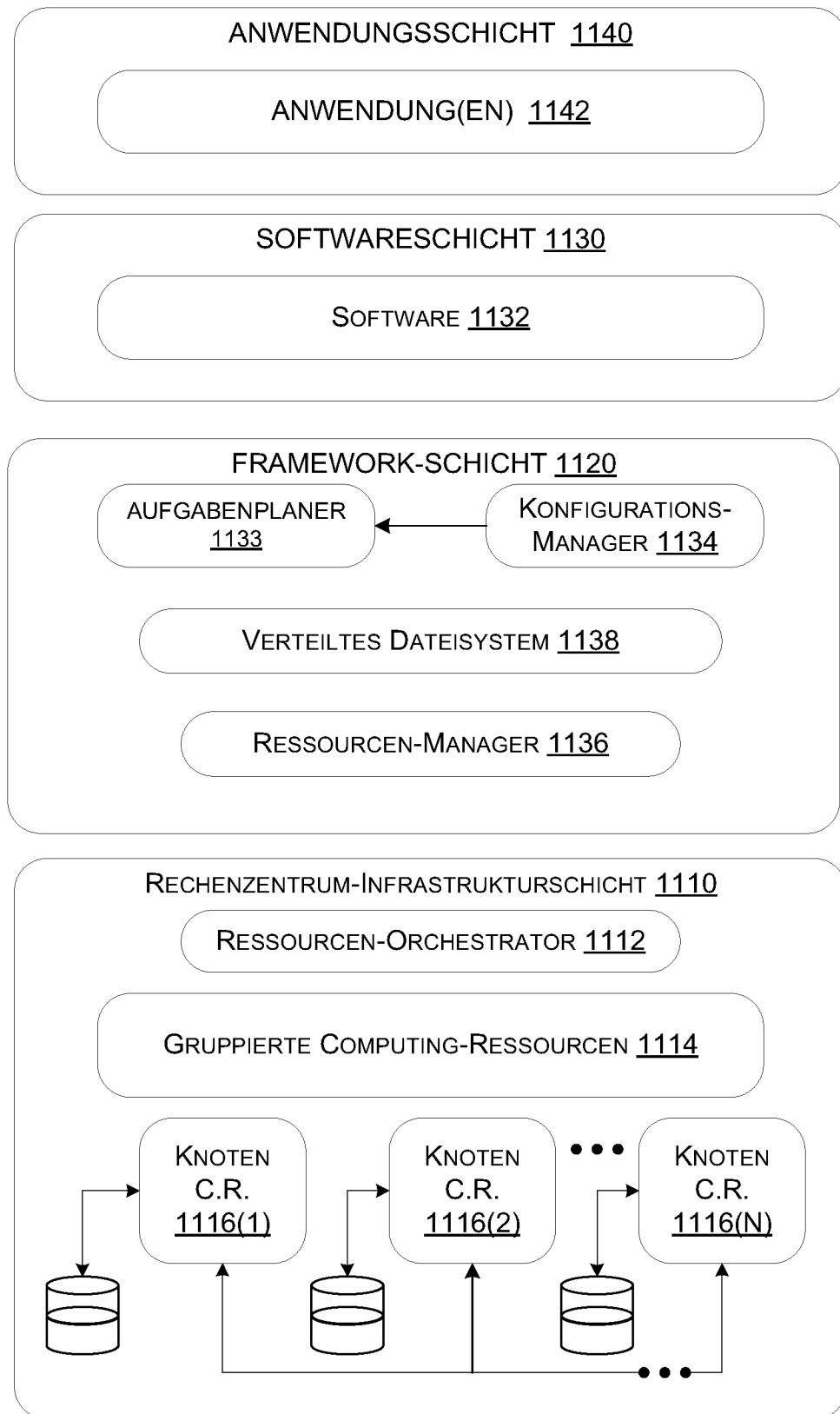


FIGUR 9D



FIGUR 10

1100
↓



FIGUR 11