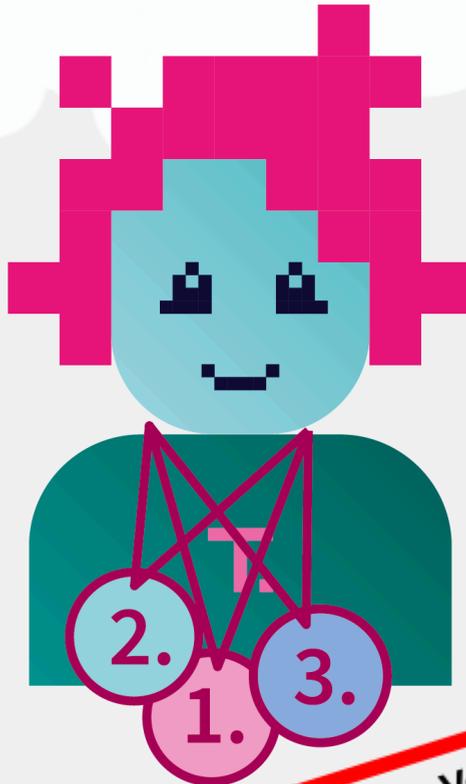


Zentrale Konzepte Großer Sprachmodelle

TINAs Lernreise



Temporäre Vorabversion!

Impressum

Große Sprachmodelle (LLMs) sind Computersysteme, die Sprache verarbeiten: Sie führen Gespräche, beantworten Fragen und erledigen Aufgaben wie Übersetzen und Zusammenfassen. Am Beispiel des fiktiven Sprachmodells TINA, einer "neugierigen KI", die lernt, menschliche Sprache zu verstehen und zu benutzen, erfährst Du, wie solche Modelle entstehen.

Diese Broschüre ist Dein Reisebegleiter, um TINA auf ihrer Lernreise zu begleiten.

Sie richtet sich an Schüler und Lehrer und zeigt auf einfache und spielerische Weise, wie große Sprachmodelle funktionieren.

Alle Inhalte sind zum Lernen gedacht und sollen Neugier auf Künstliche Intelligenz und Sprache wecken.

Autoren:

Lana Issa - lane.issa@fu-berlin.de

Ralf Romeike - ralf.romeike@fu-berlin.de

Didaktik der Informatik, Freie Universität Berlin

Königin-Luise-Str. 24-26

14195 Berlin

Deutschland

<https://ddi.mi.fu-berlin.de>

<https://computingeducation.de>



ComputingEducation

Computer science explained!

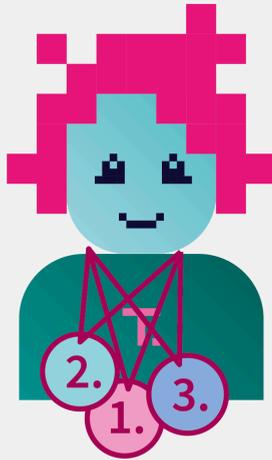
Was ist TINA?

TINA ist eine spielerische Metapher für ein großes Sprachmodell (LLM) in der Ausbildung. Auf ihrer Lernreise führt sie uns durch die wichtigsten Schritte und Phasen, die echte LLMs durchlaufen, bevor sie zu den leistungsstarken Systemen werden, die wir heute nutzen.

Ihre Geschichte ist vom echten wissenschaftlichen Trainingsprozess inspiriert, aber so erzählt, dass man leicht mitkommt: mit Bildern und Metaphern, einfachen Erklärungen und Beispielen aus dem Alltag. Schritt für Schritt lernt TINA, Texte zu zerlegen, Muster zu erkennen und am Ende verantwortungsvoll und sicher Entscheidungen zu treffen.

Es geht nicht nur darum, *wie* LLMs funktionieren. TINAs Lernreise hebt die zentralen Ideen und Konzepte dahinter hervor und macht daraus Lernaktivitäten, die nicht nur das Geheimnis lüften, wie Große Sprachmodelle entstehen, sondern auch helfen, Phänomene in der Nutzung großer Sprachmodelle zu verstehen, in dem die Grundlagen der Sprachverarbeitung mit dem Computer kennengelernt werden. Unterwegs entdecken Lernende wichtige Konzepte des maschinellen Lernens (z. B. überwacht Lernen und bestärkendes Lernen (*Reinforcement Learning*)) – und bekommen praktische Tipps, wie man Große Sprachmodelle klug nutzt und ihre Antworten kritisch hinterfragt.

TINA's Lernreise



Das ist TINA – eine neu entwickelte Künstliche Intelligenz, basierend auf einem Großen Sprachmodell. Sie beginnt gerade ihr Training – Schritt für Schritt –, um die menschliche Sprache zu verstehen und zu nutzen, damit sie Menschen helfen kann.

- Zuerst lernt sie die Muster der menschlichen Sprache, damit sie Texte lesen und sinnvoll fortsetzen kann.
- Als Nächstes lernt sie, diese Fähigkeit für bestimmte Aufgaben zu nutzen.
- Zum Schluss lernt sie, verantwortungsvoll zu entscheiden – freundlich, fair und respektvoll –, damit sie auf die richtige Weise hilft.

Wenn du TINAs Weg folgst, verstehst du nicht nur die technischen Phasen des LLM-Trainings, sondern siehst auch die menschlichen Überlegungen, die bestimmen, wie KI mit der Welt umgeht.

Während TINA Schritt für Schritt lernt, verändert sich ihr Aussehen und zeigt ihren Fortschritt. So können wir uns vorstellen, wie aus einem ganz einfachen System nach und nach eine fähige und verlässliche Assistentin wird.



Phase 1: Pre-Training (Sprechen lernen)

Am Anfang ist TINA eine unscharfe, pixelige Figur, die nicht mit Sprache umgehen kann. Sie erkennt bald Wörter als kleine Texteinheiten („Tokens“), versteht ihre Bedeutung aber noch nicht. Wenn sie Embeddings (ihre Wörter-Landkarte) und Attention (Kontext-Aufmerksamkeit) lernt, wird ihr Bild nach und nach klarer und schärfer.



Phase 2: Fine-Tuning (Lernen, Dinge zu tun)

TINAs Avatar wirkt selbstsicherer. Anhand vieler Beispiele erwirbt sie Fertigkeiten und Fachwissen, sodass sie zu bestimmten Themen arbeiten kann – z. B. Naturwissenschaft, Geschichte oder Schulthemen.



Phase 3: Alignment (Dialoge verbessern)

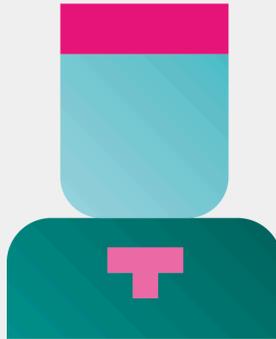
TINAs Avatar ist jetzt vollständig und schaut freundlich. Er zeigt Symbole für Empathie, Fairness und Verantwortung – ein Zeichen dafür, dass TINA sicher und angemessen mit Menschen sprechen kann.

Phase 1: Sprechen lernen

Es geht um Quantität!

Pre-Training

Vorhersage von nächsten Wörtern üben, mit sehr vielen Texten.



Wie funktioniert die Sprache eigentlich? TINA betrachtet eine große Anzahl von Texten und lernt, aus den gegebenen Wörtern das nächste Wort vorherzusagen. Dabei lernt sie, dass Wörter unterschiedliche Bedeutungen haben können. Sie lernt, Wörter je nach Kontext richtig einzuordnen, relevante Stellen im Satz zu erkennen und die Reihenfolge der Wörter zu beachten..



Zunächst sieht TINA Text nur als eine lange Folge von Zeichen, was es ihr schwer macht, ihn zu verstehen.

1.1 Tokenization

TINA baut ihren Wortschatz auf



Erinnerst du dich daran, wie du im Englisch- oder Spanischunterricht neue Wörter gelernt hast? Bevor man ganze Sätze bildet, lernt man erst viele einzelne Vokabeln und erkennt sie als Teil der Sprache. So ähnlich lernt TINA. Sie schaut sich sehr viele Texte an und zerlegt sie in ihre kleinsten Bausteine – sogenannte Tokens.

Was ist ein Token?

Ein Token ist wie die Rohfassung eines Wortes: Manchmal ist es das ganze Wort, manchmal nur ein Stück davon (z. B. „Lern“ und „-ender“ in „Lernender“). So kann TINA besser mit den vielen Wortformen umgehen, wie Lernender, Lernerfolg, Lernhilfe usw..

Warum macht man das?

Computer verarbeiten Text am besten, wenn man ihn später in Zahlen umwandeln kann – und jedem Token kann man eine eindeutige

Nummer geben. Tokenisierung bedeutet: Sätze werden in eine Folge von Tokens „geschnitten“.

Und wie geht es danach weiter?

Nach dem Zerschneiden in Tokens beginnt das eigentliche Training: TINA liest sehr viele Sätze und bemerkt Muster – welche Tokens oft vorkommen, wie sie in Sätzen zusammenpassen und wo sie wieder auftauchen. So sammelt TINA Tokens als ersten Schritt ihrer Sprachreise. Aus diesen Bausteinen kann sie Wörter (auch neue Wortformen) zusammensetzen und später Texte verstehen und erzeugen.

Aktivität 1: Tokenisieren mit TINA!

Worum geht's? Sprachmodelle wie TINA Text verarbeiten können, zerlegen sie ihn in Tokens – die kleinsten Einheiten, mit denen sie arbeiten. Diese Tokens können ganze Wörter, Wortteile oder Satzzeichen sein. In dieser Aktivität nutzen die Schüler ein online GPT-Tokenizer-Tool, um auszuprobieren, wie Computer Sätze in Tokens aufteilen, die Ergebnisse mit den eigenen Vermutungen zu vergleichen und darüber zu sprechen, was Tokenisierung für das KI-Verständnis bedeutet.

Das nimmst du mit:

- Tokenisierung ist der erste Schritt, wie ein LLM Text „liest“.
- Der gleiche Satz kann auf unerwartete Weise in Tokens zerlegt werden.
- Tokenisierung beeinflusst, wie gut ein LLM Text versteht und erzeugt.

Vorbereitung:

- Ein Internetgerät pro Gruppe oder Paar.
- Das GPT-Tokenizer-Tool:
<https://platform.openai.com/tokenizer>
- Ein Arbeitsblatt zum Eintragen von Token-Zahlen und kurzen Reflexionen.

Kurzbeschreibung:

Die Schüler testen, wie GPT Text in Tokens aufteilt, notieren die Ergebnisse und reflektieren die Muster, die ihnen auffallen.

Kontext

TINA ist ein Computersystem, das eine Sprache lernen will! So wie wir beim Lesenlernen nicht sofort ganze Sätze verstehen, sondern erst kleinere Teile erkennen, zerlegt auch TINA Sätze in kleine Stücke – die Tokens. Hilf TINA, Sätze in Tokens zu zerlegen, damit sie Sprache wie ein LLM verarbeiten kann!

Beschreibung der Aktivität

Schritt 1 – Rate die Token

Hier ist der erste Satz (in Englisch):

The happy cat jumped over the tall fence

Schätze: Wie viele Tokens macht der GPT-Tokenizer daraus? Ihr dürft nach Wörtern, Wortteilen oder Satzzeichen schätzen.

Vermutete Anzahl an Token: _____

Schritt 2 – GPT-Tokenizer nutzen

Gebt den Satz in das Tool ein (beliebiges Modell, z. B. GPT-4o) und notiert

Tatsächliche Anzahl an Token gemäß Tokenizer: _____

Hier das Ergebnis:

The screenshot shows a tokenization tool interface. At the top, it displays 'Tokens: 8' and 'Characters: 40'. Below this, the sentence 'The happy cat jumped over the tall fence' is shown with individual words highlighted in different colors. At the bottom, there are two buttons: 'Text' (which is active) and 'Token IDs'.

Schritt 3 – Mit drei Sätzen wiederholen

Schreibe auf für jeden Satz:

- Die geschätzte Anzahl der Token.
- Die Anzahl der Token im Tokenizer-Tool

Satz	Geschätzte Tokenzahl	Tokenzahl im Tool
<i>I'm learning how AI works!</i>		
<i>Die fröhliche Katze sprang über den hohen Zaun.</i>		
<i>Photosynthese ist der Weg, wie Pflanzen Nahrung herstellen.</i>		

Reflexion

Waren die Tokens immer genau die Wörter?

Nein. Manchmal ist ein ganzes Wort ein Token, manchmal wird ein Wort in mehrere Stücke geteilt.

Warum beeinflusst Tokenisierung, wie TINA Text versteht?

Sie zerlegt Wörter in nützliche Teile. So kann TINA Muster erkennen, auch bei neuen oder langen Wörtern.

Ist das so wie mit dem Vokabular, das ein Mensch hat?

Betrachten wir hierzu eine Schätzfrage: Wie viele Wörter kennst du ungefähr in deiner Sprache?

- ca. 2.500
- ca. 25.000
- ca. 250.000

Ein 12-jähriges Kind versteht im Durchschnitt etwa 25.000 Wörter (rezeptiv). Erwachsene verstehen deutlich mehr. Ein Sprachmodell wie TINA hat kein klassisches Wörterbuch, sondern ein Token-Vokabular (bis zu rund 200.000 Token). Diese Stücke sind oft Wortteile, die sich zu ganzen Wörtern und Sätzen zusammensetzen lassen. Das „Alphabet“ aus Token-Stücken wurde aus häufig vorkommenden Zeichenfolgen gelernt und deckt viele Sprachen ab.

Weil sehr viel Trainingstext Englisch ist, sind englische Wörter oft ein einzelnes Token. Deutsche Wörter – besonders lange Zusammensetzungen – werden dagegen häufiger in mehrere Tokens aufgeteilt. Englisch ist deshalb oft kompakter; Deutsch braucht für denselben Inhalt mehr Tokens. Und dann sind da natürlich noch die Besonderheiten der deutschen Sprache mit langen

Wortzusammensetzungskettenlängenerscheinungen



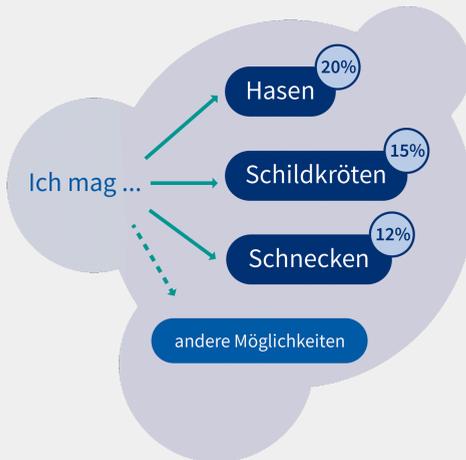
TINA kann jetzt verschiedene Tokens erkennen und aus diesen Bausteinen Wörter und Sätze bilden.



Auch wenn TINA Tokens voneinander trennen kann, weiß sie noch nicht, was sie bedeuten.

1.2 Word Prediction

TINA lernt, welches Wort wahrscheinlich als nächstes kommt.



Kennst du das, wenn du den Satz eines Freundes schon beim Sprechen selbst beenden kannst?

Sagt jemand „Es war ...“, denkst du sofort an „einmal“. Und bei „Die Katze saß auf der ...“ kommt vielen „Matte“ in den Kopf.

Genau das übt TINA auch. Sie liest sehr viele Sätze und merkt sich, welche Wörter nach anderen häufig folgen und welcher Kontext (die Wörter davor) dazu passt.

Diese Fähigkeit, das nächste Wort zu erraten, ist eine von TINAs wichtigsten Grundfähigkeiten. Sie ist nicht nur ein einzelner Schritt, sondern der rote Faden durch alle Lernphasen: Mit jeder Phase wird diese Fertigkeit genauer, damit TINA Texte schreibt, die sinnvoll, im Zusammenhang bleibend und für Menschen nützlich sind.



TINA hat nun ein “Gefühl” dafür, welches Wort als Nächstes passen könnte.



TINA kann aus den vorherigen Wörtern gute Vorhersagen machen – versteht aber die volle Bedeutung und die tieferen Zusammenhänge zwischen Wörtern noch nicht.

Aktivität: TINA und Grimms Märchen!

Basierend auf Olari, V., Zoppke, T. und Romeike, R. (2023): Sprachmodelle unplugged mit Grimms Märchen.

Worum geht's? Computersysteme wie TINA erfinden Sätze nicht völlig neu. Sie raten das nächste Wort, indem sie aus vielen Beispielen lernen, welche Wörter häufig aufeinander folgen. In dieser Aktivität bauen die Schüler ihre eigene Wort-Vorhersage-Maschine mit Sätzen aus Grimms Märchen und erleben, wie „neue“ Geschichten aus alten Mustern entstehen.

Das nimmst du mit:

- Wortvorhersage ist eine der wichtigsten Fähigkeiten eines LLM.
- Computer nutzen Wahrscheinlichkeiten und Muster, kein echtes „Verstehen“.
- Neue Texte entstehen durch Wiederverwenden und Neukombinieren von Wörtern/Token.

Vorbereitung:

- Ein Satzset aus kurzen Märchensätzen (z. B. „Der Wolf traf Rotkäppchen.“, „Die Prinzessin berührte die Spindel.“, „Es war einmal ein König.“, „Die Hexe wohnte im Wald.“, „Der Jäger öffnete die Tür.“) – als Streifen ausgeschnitten.
- Wortkarten oder weitere Satzteil-Karten (zum Mischen und Neuordnen).
- Stifte und Papier (für die entstandenen Mini-Märchen).

Kurzbeschreibung: Die Schüler zerlegen kurze Märchensätze in einzelne Wörter und ordnen sie neu, um eigene „Vorhersage-Sätze“ zu bauen. Dabei erleben sie, wie durch das Wiederverwenden und Kombinieren bekannter Wörter neue Texte entstehen. So verstehen sie spielerisch, dass TINA beim Schreiben keine echten Ideen hat, sondern Wahrscheinlichkeiten nutzt, um passende nächste Wörter auszuwählen.

Kontext

TINA lernt, wie man Geschichten erzählt! Statt eigene Ideen zu erfinden, schaut sie sich zuerst Märchen an – zum Beispiel die der Brüder Grimm. Sie achtet darauf, welche Wörter oft aufeinander folgen: Auf „Es war einmal ...“ kommt meistens „... ein“ oder „... eine“. Wenn man solche Wortpaare aneinanderreicht, kann TINA neue Märchen „schreiben“, die vertraut klingen – und manchmal auch sehr witzig sind!

Beschreibung der Aktivität

Schaut euch die folgenden Sätze aus Grimms Märchen an:

“Es war einmal ein Königssohn, der bekam Lust, in der Welt umherzuziehen”

“Ein Vater hatte zwei Söhne, davon war der älteste, klug und geschickt.”

“Es war ein armer Bauer, der saß Abends beim Herd und schürte das Feuer.”

“Es war einmal ein armes, frommes Mädchen, das lebte mit seiner Mutter allein”

“Vor Zeiten war ein Schneider, der drei Söhne hatte und nur eine einzige Ziege.”

“Es war ein armer Mann und eine arme Frau, die hatten nichts als eine kleine Hütte.”

1. **Text tokenisieren:** Schneide die Sätze in Wort-Karten.

“Es war einmal ein Königssohn, der bekam Lust, in der Welt umherzuziehen”

[Es] [war] [einmal] [ein] [Königssohn] [,] [der] [bekam] [Lust] [,]
[in] [der] [Welt] [umherzuziehen]

2. **Startwort wählen:** Beginne mit einem typischen Anfang, z. B. „Der/Die/Das“ oder „Es“ (wie in „Es war ...“).

Es _____

3. **Nächstes Wort raten:** Schau in die gegebenen Sätze und entscheidet, welches Wort meistens folgt. (Gibt es mehrere Möglichkeiten? Dann per Würfel entscheiden.).

“Der” → *“bekam”, “saß”, “drei”*

“Es” → *“war”*

“war” → *“ein”, “einmal”*

“armer” → *“Bauer”, “mann”*

“frommes” → *“Mädchen”*

“drei” → *“Söhne”*

4. **Satz bauen:** Lege ein Token nach dem anderen an. Nutze immer das zuletzt gelegte Wort, um das nächste auszuwählen.

“Es” → “war”

“war” → “einmal”

“einmal” → “ein”

“ein” → “armer”

“armer” → “Mann”

.....

Ergebnis: *“Es war einmal ein armer Mann, der saß Abends beim Herd und bekam Lust, in der Welt umherzuziehen.”*

5. **Weiterbauen:** Füge mehr Wörter hinzu, bis ein Punkt kommt oder die Länge passt – fertig ist Dein Mini-Märchen!

Reflexion

Warum klingen manche Sätze lustig, aber trotzdem wie ein Märchen?

- Weil wir bekannte Märchen-Wörter und Satzanfänge benutzen („Es war einmal...“, „die Prinzessin...“).
- Die Wörter werden aber nur nach Häufigkeit aneinandergehängt – ohne echtes Verstehen. Deshalb passen sie manchmal nicht perfekt zusammen (Grammatik, Sinn, Reihenfolge) und wirken komisch.
- Trotzdem erinnert der Stil an Märchen (Figuren, Orte, typische Wendungen), daher klingt es „märchenhaft“.

Was zeigt das darüber, wie TINA anfängt zu lernen?

- TINA lernt zuerst Muster: welches Wort oft auf welches folgt.
- Sie kann damit Sätze fortsetzen, ohne die Bedeutung wirklich zu kennen.
- Für besseren Zusammenhang braucht TINA später mehr Kontext und weitere Techniken (z. B. die „Wörter-Landkarte“/Embeddings und Aufmerksamkeit/Attention).

Weitere Aktivitäten

Sprachmodelle unplugged mit Grimms Märchen:

Unplugged-Aktivität, in der grundlegende Konzepte großer Sprachmodelle anhand eines didaktisch reduzierten, auf einer Markow-Kette basierenden Modells vermittelt werden. Als Text-Korpus werden Sätze aus Märchen der Gebrüder Grimm verwendet.

<https://dl.qi.de/items/44a62375-bd2e-4a81-bb59-8f6595e117de>

Auf die Idee aufbauend, wurden die folgenden Implementationen entwickelt:

https://www.kiki-labor.fau.de/files/2024/10/20_Infotexte_LY_Maerchenstunde.pdf

SOEKIAGpt

<https://www.soekia.ch/gpt.html>

https://www.soekia.ch/IBiS_1-1_PB-2_Hielscher_SoekiaGPT.pdf

Snap!GPT:

<https://www.informatischebildung.de/index.php/ibis/article/view/31>

1.3 Embedding

TINA gruppiert Wörter mit ähnlicher Bedeutung.



Lass uns ein Spiel spielen: Welche drei Wörter gehören zusammen – und welches passt nicht dazu?

Hund | Banane | Katze

Deutschland | Frankreich | Boot

Stuhl | grün | rot

Im ersten Beispiel merkst du schnell: Hund und Katze gehören zusammen, Banane nicht – wir Menschen verstehen die Bedeutungen (beides Tiere) und können Wörter natürlich gruppieren. Stell dir vor,

du bist in China, kannst kein Chinesisch lesen, hast aber ein fotografisches Gedächtnis: Du würdest trotzdem zuordnen lernen, welche Zeichen du eher auf Speisekarten siehst und welche auf Verkehrsschildern. So ähnlich geht es TINA:

TINA versteht die Bedeutung von Wörtern nicht wie wir. Stattdessen liest sie sehr viele Texte und merkt: Welche Wörter kommen häufig in ähnlichen Sätzen vor? Wörter, die oft in ähnlichen Zusammenhängen (Kontexten) auftauchen, werden in TINAs „Kopf“ zu Nachbarn – zum Beispiel liegt „Hund“ nahe bei „Katze“, aber weit weg von „Banane“.

So entsteht eine Wörter-Landkarte: Nähe bedeutet Ähnlichkeit in der Verwendung. Mit dieser Landkarte kann TINA Wörter wählen, die zum Kontext passen – dadurch bleiben ihre Sätze zusammenhängend und stimmig.

Aktivität 3: Hilf TINA, eine Wörter-Landkarte zu bauen!

Worum geht's? Computer verstehen Wörter nicht eins nach dem anderen. Stattdessen lernen sie aus dem Umfeld, in dem Wörter auftauchen. In dieser Aktivität ordnen die Schüler Wörter räumlich an: Nah beieinander, wenn sie oft in ähnlichen Situationen vorkommen – so entsteht eine Wörter-Landkarte.

Das nimmst du mit:

- Wörter bekommen Bedeutung durch ihren Gebrauch mit anderen Wörtern.
- Ein Wort kann zu mehreren Bereichen gehören – je nach Kontext.
- Diese Verbindungen helfen TINA, das nächste Wort in einem Satz zu erraten

Vorbereitung:

- Karten für einzelne Wörter, z. B. Hund, Katze, Tierarzt, Leine, Bank, Geld, Park
- Großes Papier/Plakat oder Tafel/Whiteboard
- Stifte/Marker, ggf. Klebepunkte/Haftnotizen.

Kurzbeschreibung:

Die Schüler erstellen eine Wörter-Landkarte, indem sie Wörter zusammenlegen, die im gleichen Satz vorkommen. So sehen sie, wie diese „Karte“ TINA hilft, das nächste Wort in einem Satz vorherzusagen.

Kontext

TINA möchte eine eigene Wörter-Landkarte aus Substantiven bauen (wir beschränken uns an dieser Stelle zur Vereinfachung auf Substantive, das gleiche Prinzip gilt für alle Wörter). Sie versteht Wörter nicht wie ein Mensch, aber sie bemerkt, welche Wörter oft zusammen vorkommen. Wenn Hund oft mit Leine steht, liegen diese Wörter nah beieinander. Bank ist knifflig: Neben "Geld" meint sie das Geldinstitut, neben "Park" eher die Sitzbank. Ein Wort kann also mehrere Nachbarschaften haben – Kontext entscheidet.

Beschreibung der Aktivität

Schritt 1: Wörter-Landkarte anlegen

„Die **Hunde** jagten die **Katzen** im **Park**.“

„Der **Prinz** und die **Prinzessin** feierten mit einer großen **Torte**.“

„Sie ging zur **Bank**, um ihr **Geld** einzuzahlen.“

„Im **Park** stand eine große **Bank**“

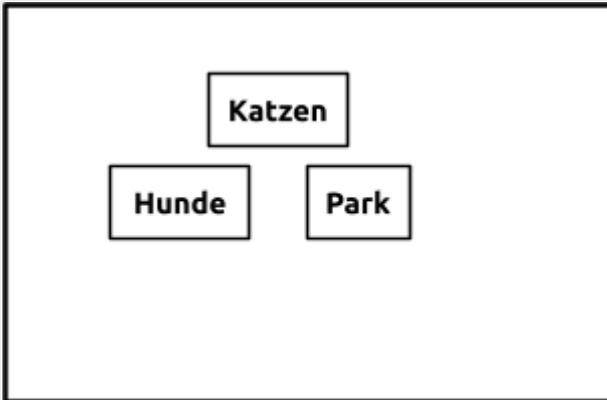
„Die **Katzen** jagten **Kaninchen** und **Vögel**.“

Schreibe jedes Substantiv der gegebenen Sätze auf eine eigene Wortkarte.

Bilde nun Gruppen/Cluster: Welche Wörter passen zusammen, weil sie häufig in ähnlichen Sätzen vorkommen? Folge dem Beispiel:

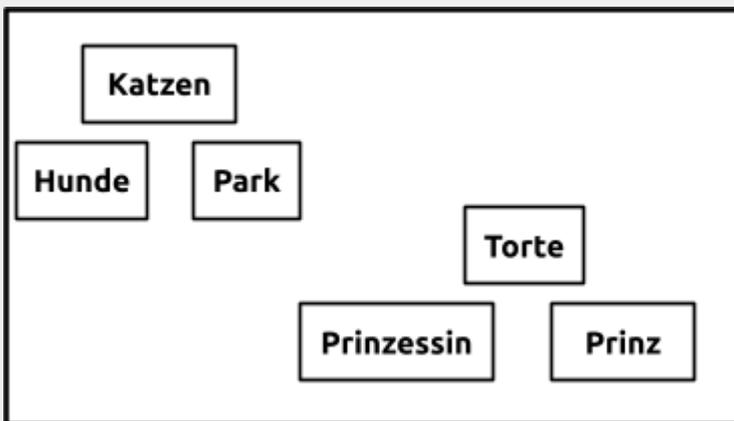
„Die **Hunde** jagten die **Katzen** im **Park**.“

„Hunde“, „Katzen“ und „Park“ legen wir nah beieinander.



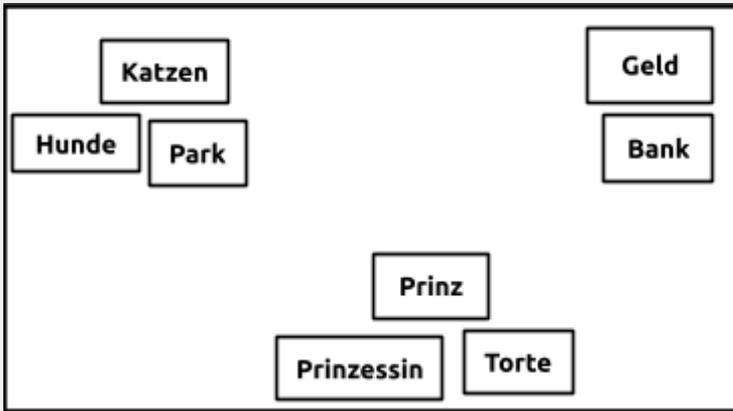
„Der **Prinz** und die **Prinzessin** feierten mit einer großen **Torte**.“

Beim zweiten Satz kommen „Prinz“, „Prinzessin“ und „Kuchen“ dazu.



„Sie ging zur **Bank**, um ihr **Geld** einzuzahlen.“

Jetzt tauchen neue Wörter auf: „**Geld**“ und „**Bank**“:

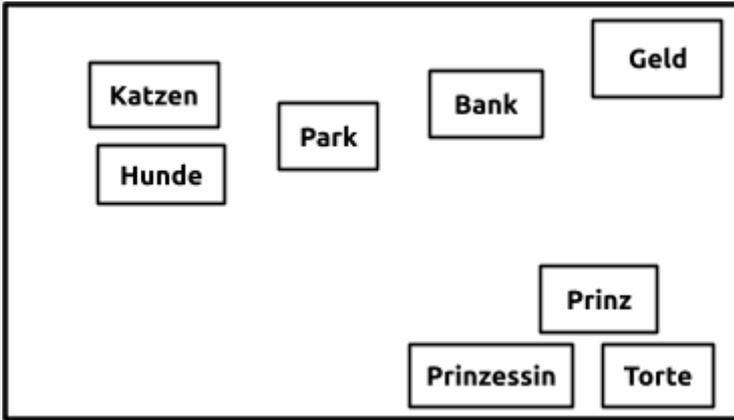


„Im **Park** stand eine große **Bank**“

In diesem Satz lernen wir keine neuen Wörter, aber wir sehen bekannte Wörter in neuer Verwendung.

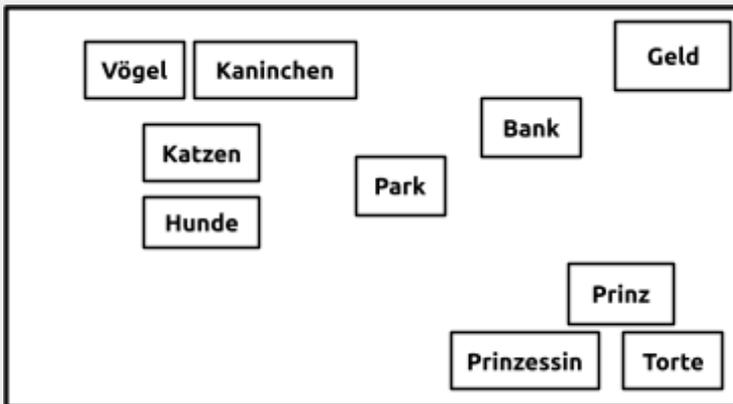
Im Beispiel taucht „Bank“ zusammen mit „Park“ auf. Bisher verband TINA „Bank“ nur mit „Geld“. Jetzt lernt sie die Bedeutung als Sitzbank im Park kennen.

Darum rücken wir auf TINAs Wörter-Landkarte Bank und Park näher zusammen. So wird die Verbindung zwischen beiden stärker – und „Bank“ bekommt mehr Bedeutungsschattierungen. Das hilft TINA, das Wort in Zukunft passender zu verwenden.



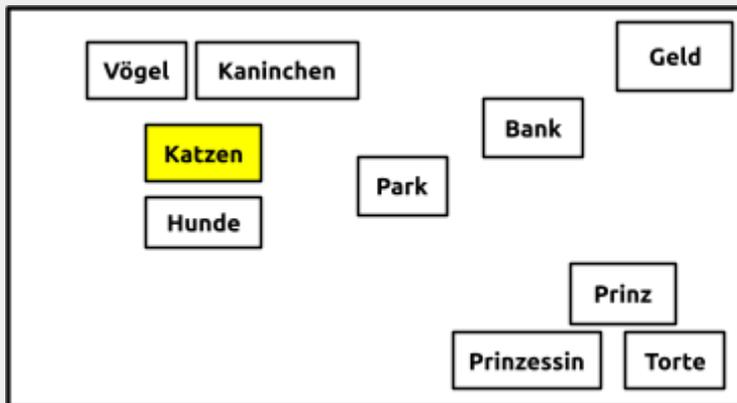
„Die *Katzen* jagten *Kaninchen* und *Vögel*.“

Im letzten Satz kommen „Kaninchen“ und „Vögel“ zusammen mit „Katze“ vor – also legen wir „Kaninchen“ und „Vögel“ nah an „Katze“ auf die Wörter-Landkarte.



Schritt 2: Wörter-Landkarte anwenden

Zufallswort wählen: Bestimme ein beliebiges Wort von der Wörter-Landkarte, z. B. „Katzen“.

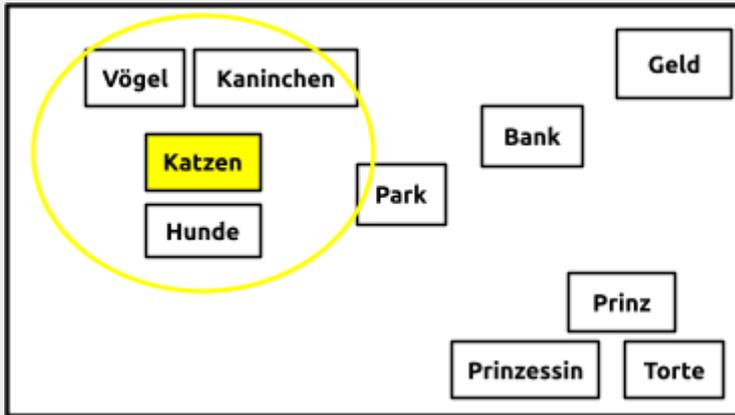


Nachbarn finden: Schau auf der Karte, welche Wörter nah bei "Katzen" liegen. Zur Erinnerung: Das sind Wörter, die oft in Sätzen gemeinsam vorkommen.

Smalltalk machen: TINA soll nun die Wörter-Landkarte anwenden, um freundlich nachzufragen:

„Ich glaube, du magst Katzen, oder? Magst du auch X, Y und Z?“

Zeichne einen Kreis um „Katzen“ und vergrößere ihn langsam, bis mindestens drei weitere Wörter im Kreis liegen. Wähle drei Wörter als X, Y und Z.



Setze nun die Wörter ein, z. B.:

„Ich glaube, du magst Katzen, oder? Magst du auch Vögel, Kaninchen und Hunde?“

Schritt 3: Eigene Wörter-Landkarte bauen und nutzen

Wähle nun einen eigenen Text (z. B. ein Märchen) und wende die Methode an:

- **Text auswählen & zerlegen:** Markiere wichtige Wörter/Begriffe, schreibe sie auf Wortkarten.
- **Landkarte legen:** Ordne die Karten so an, dass ähnlich verwendete Wörter nah beieinander liegen.
- **Verbindungen einzeichnen:** Ziehe Linien zwischen Wörtern, die oft zusammen vorkommen.
- **Sätze bauen:** Nutze die Landkarte, um eigene Sätze zu formulieren – wähle dafür Wörter aus derselben Nachbarschaft.

- **Kurz reflektieren:** Was liegt nah beieinander? Gab es Grenzfälle (z. B. mehrdeutige Wörter)? Wie hat Dir die Karte beim Formulieren geholfen?

Reflexion

Was sagt Dir das darüber, wie TINA „Sprache versteht“?

- TINA achtet vor allem auf Muster und Nachbarschaften von Wörtern – nicht auf „echte“ Bedeutungen.
- Sie benutzt ihre Wörter-Landkarte, um Wörter zu finden, die oft zusammen vorkommen oder ähnlich verwendet werden.
- Sie kann erkennen, welche Wörter zusammenpassen, versteht aber Grammatik und Satzbau noch nicht wirklich.
- Deshalb klingen manche Sätze stimmig, andere aber holprig oder komisch.



TINA hat jetzt eine “Wörter-Landkarte”, die zeigt, welche Wörter in Gebrauch und Bedeutung nah beieinander liegen.



TINA kann sehen, welche Wörter zusammengehören, versteht aber Grammatik und Satzstruktur noch nicht.

1.4 Positional Encoding

TINA lernt, die Stellung der Wörter zu erkennen und zu berücksichtigen.



Schon kleine Unterschiede in der Wortreihenfolge können die Bedeutung stark verändern:

Erkennst du den Unterschied zwischen „Peter hat nur die Aufgabe gelöst.“ und „Nur Peter hat die Aufgabe gelöst.“?

Damit TINA solche Unterschiede berücksichtigen kann, merkt sie sich nicht nur die Wörter, sondern auch ihre Position im Satz. Dazu bekommt jedes Wort eine Art „Positionsnummer“ (1, 2, 3 ...). Aus sehr

vielen Beispielen lernt TINA, welche Reihenfolgen typisch sind – und wann eine andere Reihenfolge etwas anderes bedeutet. So kann sie nach und nach die passende Wortreihenfolge wählen.



TINA weiß jetzt: Die Position eines Wortes kann die Bedeutung eines Satzes verändern.



TINA berücksichtigt Wörter und ihre Reihenfolge, braucht aber noch Übung, um längere Gespräche flüssig und zusammenhängend zu führen.

1.5 Attention

TINA lernt, mithilfe des Kontexts relevante Stellen hervorzuheben und Mehrdeutigkeiten aufzulösen.



Nachdem TINA gelernt hat, dass die Position von Wörtern die Satzbedeutung beeinflusst, muss sie noch erkennen, welche Teile eines Satzes gerade wichtig sind. Dabei hilft Attention. Stell dir einen Scheinwerfer vor: TINA leuchtet damit die relevanten Wörter und Positionen an, um die Fortsetzung besser vorherzusagen.

Beispiel: „Die Katze sprang auf die Couch, weil sie weich war.“

Mit Attention merkt TINA: „sie“ bezieht sich auf die Couch (und nicht auf die Katze). So sieht sie nicht nur einzelne Wörter, sondern auch Verbindungen zwischen Wörtern – sogar wenn sie weit auseinanderstehen.

Manchmal schaut TINA auf alle Wörter im Satz gleichzeitig, um Zusammenhänge zu erkennen. In anderen Situationen beachtet sie nur die Wörter, die sie schon geschrieben hat, um zu entscheiden, welches Wort als Nächstes kommt. In beiden Fällen hilft ihr Attention, fokussiert zu bleiben und sinnvolle Sätze zu bauen.



TINA kann jetzt wichtige Stellen (Wörter/Positionen) im Satz hervorheben.



Sie gewichtet Beziehungen zwischen Wörtern und erkennt, welche einander am meisten beeinflussen – braucht aber noch Übung, um immer vollständig kohärente Antworten zu geben.

Activity 4: TINA's Attention!

Worum geht's? Manche Wörter haben mehrere Bedeutungen. „Bank“ kann eine Sitzbank oder ein Geldinstitut sein, „Schloss“ ein Märchenschloss oder ein Türschloss. Für uns ist der Kontext meist klar, für Computer ist das schwierig.

In dieser Aktivität lernen die Schüler, wie TINA wichtige Wörter im Satz hervorhebt, um die richtige Bedeutung zu finden. Diesen Trick nennt man Aufmerksamkeits-Mechanismus (Attention).

Das nimmst du mit:

- Sprachmodelle müssen auf wichtige Wörter achten, um den Kontext zu erkennen.
- Hinweiswörter (z. B. „Park“, „Geld“, „König“, „Schlüssel“) helfen, Mehrdeutigkeiten zu klären.
- Attention ist ein Schlüssel, damit Sprachmodelle bessere Vorhersagen machen.

Vorbereitung:

- Stifte in vier Farben (z. B. schwarz, rot, grün, blau)
- Ausdruck eines „Attention“-Blatts mit Pfeilen (oder leere Kästchen/Zeilen, um Pfeile selbst zu zeichnen)
- Arbeitsblätter (ausgedruckt oder digital)

Kontext

„Es geschah, dass an dem Tage, wo es gerade fünfzehn Jahr alt ward, der König und die Königin nicht zu Haus waren, und das Mädchen

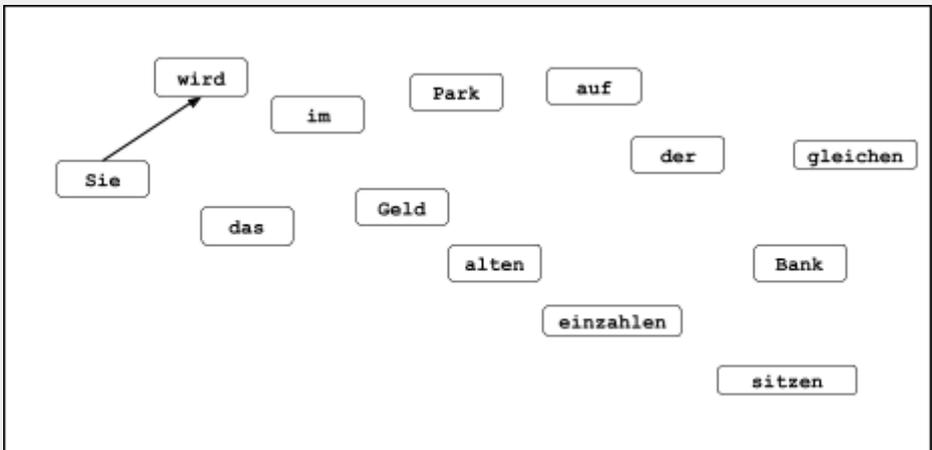
ganz allein im Schloss zurückblieb...“ und „In dem Schloss steckte ein verrosteter Schlüssel, und als es umdrehte, sprang die Türe auf...“

Für uns Menschen ist klar: Im ersten Satz steht „Schloss“ für ein märchenhaftes Gebäude, im zweiten für eine rostige Schließvorrichtung. Für einen Computer ist diese Erkennung Aufgabe jedoch alles andere als einfach. Wie findet ein Sprachmodell die wichtigen Wörter, die ihm helfen, die richtige Bedeutung zu erkennen?

Activity Description

1. Sätze speichern:

In deinem neuen Sprachmodell sind bereits einige Wörter vorhanden. Speichere nun den Satz „Sie wird im Park auf der gleichen Bank sitzen“ genau wie bei „Grimms neuen Märchen“, indem du die Wörter mit schwarzen Pfeilen verbindest.



Speichere nun noch folgende zwei Sätze in deinem Sprachmodell. Füge hierfür jeweils die Pfeile in der entsprechenden Farbe hinzu, auch dann, wenn bereits ein Pfeil in einer anderen Farbe zwischen den Wörtern vorhanden ist.

Sie wird das Geld auf der gleichen Bank einzahlen

Sie wird im Park auf der alten Bank sitzen

Sicherlich hast du festgestellt, dass alle drei Sätze auf den ersten Blick recht ähnlich sind, aber eine unterschiedliche Bedeutung haben. Man erkennt, dass viele Wörter vor „sitzen“ und „einzahlen“ fast gleich sind.

2. Das richtige Wort finden:

„Sie wird das Geld auf der alten Bank...“ TINA soll das richtige Wort finden. Mit unserem Sprachmodell ist es allerdings nicht möglich das Wort, das nach „Bank“ kommt, sicher vorherzusagen. Sogar dann, wenn wir weitere Worte aus den häufigen Vorgeschichten einbeziehen würden. Erkennst du den Grund dafür? Besprich deine Gedanken dazu mit deinem Sitznachbarn (wenn er/sie fertig ist, mit der Aufgabe 1).

Glücklicherweise können wir TINA hier mit der richtigen Antwort helfen. Speichere noch diesen vierten Satz in deinem Sprachmodell:

Sie wird das Geld auf der alten Bank _____

3. Wichtige Wörter erkennen:

Untersuche nun dein Modell. Notiere, auf welche Wörter unser Sprachmodell TINA die Aufmerksamkeit richten muss, um zu entscheiden, ob „einzahlen“ oder „sitzen“ nach „Bank“ kommt. Beschreibe, woran TINA diese besonderen Wörter erkennen kann, die

das richtige Verständnis, also den richtigen Kontext, herstellen und besprich deine Beobachtung mit deinem Sitznachbarn.

Wieso helfen die Wörter „alten“ und „gleichen“ oder Wörter wie „wird“ nicht dabei, die Bedeutung von „Bank“ zu bestimmen und das nächste Wort passend vorherzusagen? Halte deine Vermutungen dazu auf dem Blatt fest und besprich dies ggf. mit deinem Nachbarn.

Welche Farbkombination haben die Pfeile, die auf die besonderen Wörter zeigen, um die Bedeutung von „Bank“ zu bestimmen?

Sitzgelegenheit: _____

Geldinstitut: _____

Kennzeichne die besonderen Wörter im Sprachmodell durch Umkreisen mit ihren Farben jeweils für die Bedeutung als „Sitzgelegenheit“ oder als „Geldinstitut“.

4. TINA hat gelernt:

Das Sprachmodell TINA hat nun gelernt, zwei unterschiedliche Kontexte zu unterscheiden. Schauen wir uns nochmal den ersten Satz an.

Sie wird im Park auf der gleichen Bank sitzen.

Unterstreiche in dem Satz die wichtigen Wörter, die die Bedeutung von „Bank“ bestimmen mit den entsprechenden Kontextfarben aus dem Sprachmodell.

TINA kann jetzt die wichtigen Wörter erkennen, die den richtigen Kontext kennzeichnen! Durch „Aufmerksamkeit“ auf besondere

Wörter kann das Sprachmodell viel besser vorhersagen, welches Wort als nächstes kommt.

Reflection

Was sind „wichtige“ Wörter? Wodurch sind diese gekennzeichnet?

Antwort: „Wichtige“ Wörter sind die, die den Sinn eines Satzes besonders stark tragen. Sie sind oft durch ihre Rolle im Kontext gekennzeichnet, zum Beispiel Substantive oder Verben, die die Hauptidee ausdrücken.

Was bedeutet der Aufmerksamkeitsmechanismus für die Anwendung von Sprachmodellen?

Antwort: Der Aufmerksamkeit Mechanismus sorgt dafür, dass das Modell sich auf die entscheidenden Wörter konzentrieren kann, anstatt alle Wörter gleich zu behandeln. So kann man besser einschätzen, welches Wort im nächsten Schritt sinnvoll ist.

Hat das Sprachmodul dadurch ein „Verständnis“? Welche Limitierungen bleiben? Wie wirkt das auf Individuum und Gesellschaft?

Antwort: Ein Sprachmodell gewinnt dadurch kein echtes Verständnis. Es erkennt Muster und Zusammenhänge, aber ohne eigenes Bewusstsein. Limitierungen bleiben: es kann Fehler machen, Vorurteile übernehmen oder Dinge erfinden. Für das Individuum kann es hilfreich sein (z. B. beim Schreiben oder Lernen), gleichzeitig müssen Gesellschaft und Bildung kritisch damit umgehen und Grenzen kennen.

Game: Chat with TINA (part 1)

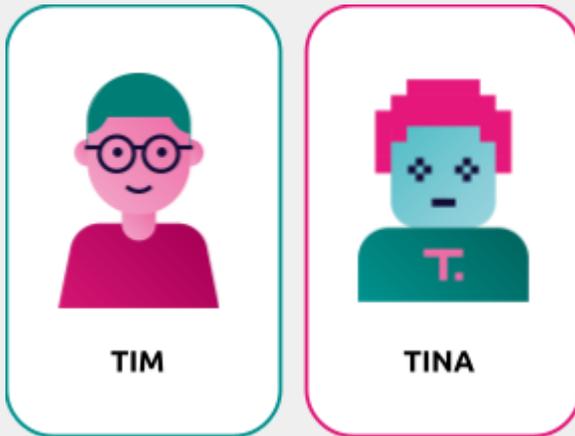
TINA has read a huge amount of text and learned how words usually go together. She can finish sentences, answer easy questions, and sound really good at talking. But **TINA** doesn't truly know facts, she's just guessing which words should come next.

Now we will play a game where you take turns being the human (**TIM**) and being the AI (**TINA**), so you can experience how TINA 'talks' and builds sentences.

Step 1: Setup your roles

Role A (**TIM**): You are the human talking to TINA. Your job is to ask TINA questions

Role B (**TINA**): You are TINA, the AI. You answer using the templates and word maps you have.



Step 2: Check out your role instructions

Instructions for player 1 (TIM)

How to speak:

- Ask short and clear questions (e.g., “Can you make a sentence with ‘dog?’”).
- Give TINA extra details to help her guess better (e.g., “Write a sentence with ‘dog’ and ‘park’”).
- Be curious, but remember: TINA doesn’t really “*understand*” she just predicts words.

Your task: Notice when TINA’s answers sound right but may not actually be correct

Examples to ask



Make a sentence with the word
dog.

Write a short story with these
words: dog, park, ball.

“The cat sat on the mat, and it
was fluffy. Who was fluffy?”

Instructions for player 2 (TINA)

How to speak:

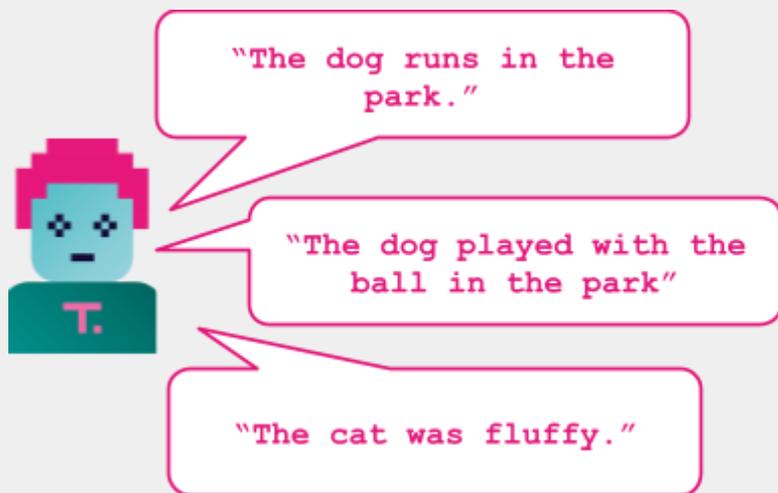
- Answer by guessing what words come next, based on patterns.
- Try to complete sentences or respond in a way that "sounds good."
- Don't explain your thinking, just give the answer.

Your limits:

- You don't really know facts.
- You might say things that sound true but aren't.

Your task: Use the given word maps or sentence starters to build responses.

Examples to answer



Reflection

1. What was easy for TINA?

- She could finish sentences and make them sound correct.
- She could combine words that fit together.
- She could answer simple, clear questions.

2. What was hard for TINA?

- Understanding the real meaning of the question.
- Knowing if her answers were actually true.
- Talking about things she has never “seen” in her training data.

3. What does this show about how AI “understands” language?

- AI does not really understand like humans.
- It just predicts words that often go together.
- It can sound smart but still make mistakes.

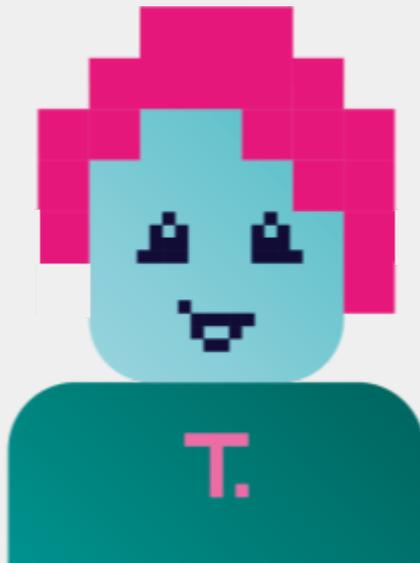
Phase 2: Lernen, Dinge zu tun

Quality matters!

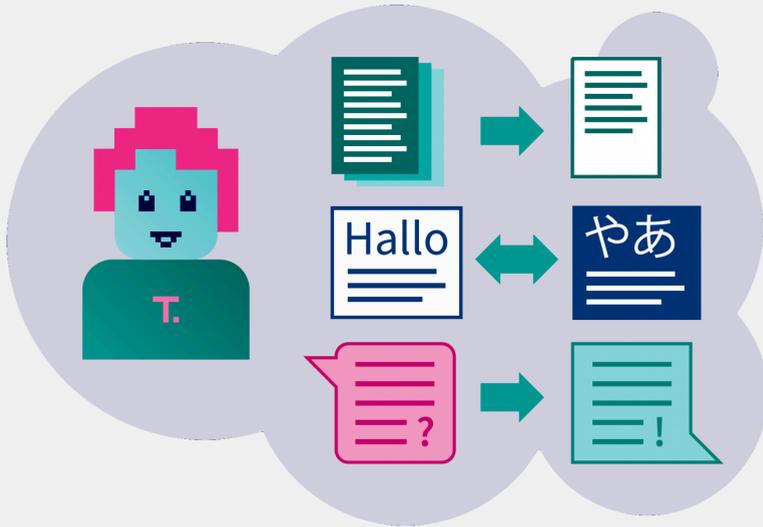
Fine-Tuning

Verschiedene Fertigkeiten entwickeln, mit ausgesuchten Datensätzen.

TINA verfeinert ihr Verständnis, indem sie sich beim Fine-Tuning auf bestimmte Themen konzentriert und mit ausgewählten Beispielen zu bestimmten Themen arbeitet, um ihr Wissen zu verfeinern und anzupassen. Auf diese Weise lernt sie, relevante Informationen auszuwählen und den Inhalt ihrer Konversation besser auszurichten.



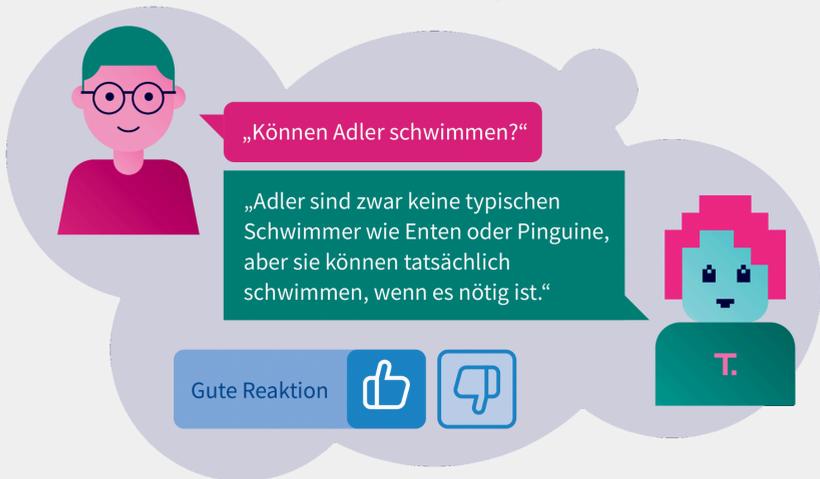
2.1 Task Tuning



In dieser Phase lernt TINA, ihr Wissen klar, hilfreich und leicht verständlich zu vermitteln. Sie übt, Informationen zu erklären, Fragen zu beantworten, Texte zusammenzufassen und Schritt-für-Schritt-Anleitungen zu geben. Durch menschliches Feedback erkennt sie, welche Antworten hilfreich, höflich und verständlich sind. Indem sie aus diesen Beispielen lernt, verbessert TINA ihre Kommunikationsfähigkeiten, sodass andere Menschen ihren Ausführungen leicht folgen, daraus lernen und sie anwenden können. Dieser Schritt verwandelt ihr Wissen in praktische Fähigkeiten und macht sie zu einer zuverlässigen und nützlichen HelferIn.



2.2 Feedback Tuning

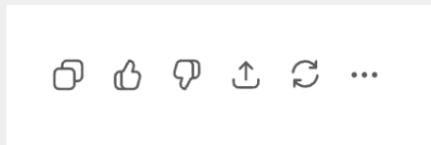


TINA lernt unterschiedliche Fertigkeiten. Am Ende wird TINA darin trainiert, verschiedene Fähigkeiten einzusetzen; zum Beispiel Geschichten erzählen, beim Lernen helfen oder höflich und freundlich mit Menschen sprechen. So kann sie ihr Wissen praktisch anwenden und in echten Gesprächen nützlich sein.

Nach Schritt 2 kann TINA nicht nur Wörter vorhersagen, sondern auch verschiedene sprachbezogene Aufgaben übernehmen, z.B. die Frage beantworten, warum ein Zeppelin fliegt oder die Kernaussagen eines Textes aufzählen.

Manchmal sind die Antworten von TINA jedoch nicht angemessen oder entsprechen nicht den Erwartungen für bestimmte Kontexte. Daher muss TINA noch lernen, ihre Textproduktion besser zu kontrollieren.

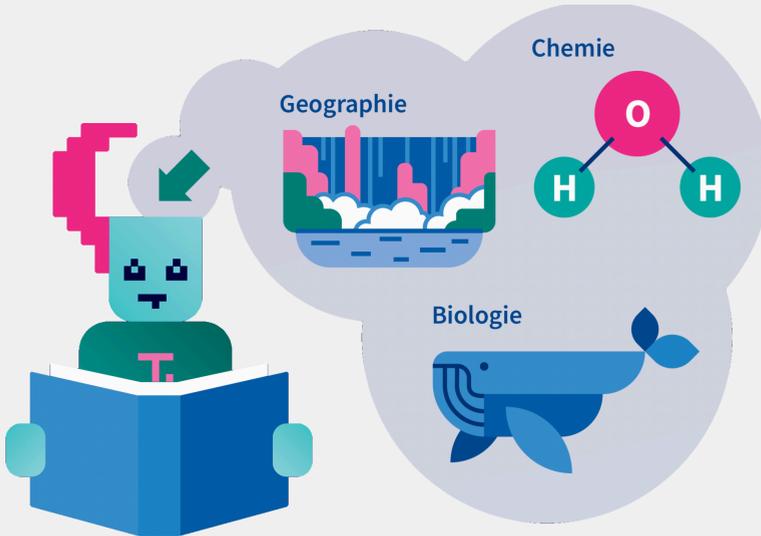
Do you notice the buttons where you can tell ChatGPT if the answer was good or not?



This is how you can contribute to the Fine Tuning of the next ChatGPT model!



2.3 Domain Tuning



In this phase, TINA goes beyond just speaking the language. She dives into special subjects like history, science, and math. By training on many texts from these areas, she learns important facts, terms, and ideas. Now she can not only chat in everyday language but also give precise answers in school subjects.



Game: Chat with TINA (part 2)

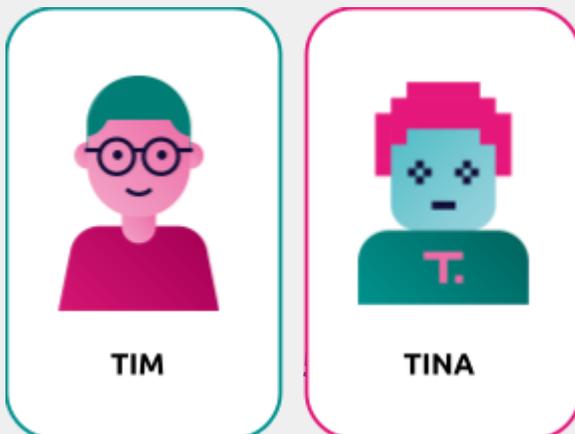
TINA has not only read a huge amount of text, but she has also practiced with examples of conversations and special topics like science, history, or math. She can now answer questions more smoothly, follow instructions better, and stay on topic more often. She can also explain ideas, summarize, or translate. But she still doesn't truly know facts, she only predicts which words should come next, and her knowledge is limited to what she has seen in training. Her answers sound more useful and complete than before, but they can still be wrong or incomplete.

Now we will play a game where you take turns being the human (**TIM**) and being the AI (**TINA**), so you can experience how TINA 'talks' and builds sentences.

Step 1: Setup your roles

Role A (**TIM**): You are the human talking to TINA. Your job is to ask TINA questions

Role B (**TINA**): You are TINA, the AI. You answer using the templates and word maps you have.



Step 2: Check out your role instructions

Instructions for player 1 (TIM)

How to speak:

- Give her instructions or small tasks (e.g., "Write two sentences with the word dog").
- Try asking step-by-step questions, because TINA is now trained to follow instructions better.
- Explore what she can do well, like giving examples, finishing stories, or answering more detailed prompts.

Your task: Notice how well TINA follows your instructions step by step, and observe how her answers improve in structure and detail, but also pay attention to areas where she might still misunderstand or give overly literal responses.

Examples to ask



"Can you tell me a funny story about a dog that talks?"

Explain step by step how to convert 45°C to Fahrenheit.

Can you translate the word "Squirrel" to German?

Instructions for player 2 (TINA)

How to speak:

- Answer step by step when prompted: If TIM asks a multi-step or detailed question, break down your answer, step by step.
- Show TINA's strength after fine-tuning, finish stories, or provide detailed explanations.
- Ask clarifying questions if needed: If TIM's instructions are unclear, ask for more information before responding.

Your limits:

- Your answers are still not very smooth and natural.
- You cannot understand hidden meanings like sarcasm or vague instructions from TIM

Examples to answer



A dog cannot talk because dogs do not have the ability to speak human language. Therefore, I cannot tell a story about a talking dog."

1. Take the temperature in Celsius: 45°C .
2. Multiply it by $9/5$: $45 \times 9/5 = 81$.
3. Add 32 to the result: $81 + 32 = 113$.
4. The temperature in Fahrenheit is 113°F .

"The German word for 'Squirrel' is 'Eichhörnchen'."

Reflection

1. What was easy for TINA?

- Following instructions step by step.
- Giving detailed examples or completing tasks clearly.
- Translating words or doing calculations correctly.

2. What was hard for TINA?

- Understanding the intent behind a question.
- Making her answers sound natural in conversation.
Deciding what is appropriate or relevant beyond the literal task.

3. What does this show about how AI “understands” language?

- AI still doesn't truly understand meaning or intent.
- It has learned patterns for following instructions and giving structured answers.
- It can seem precise and detailed but may misinterpret nuance or context.

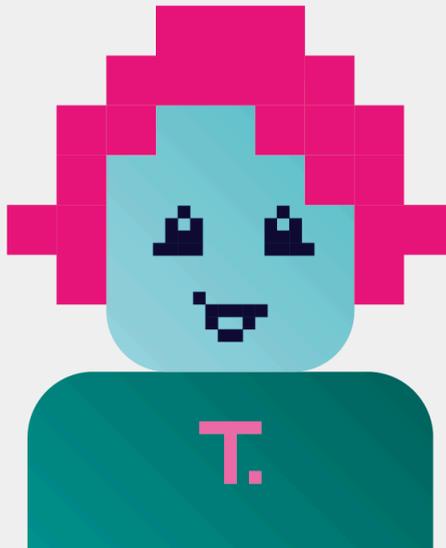
Phase 3: Dialoge verbessern

Nett Sein!

Alignment

Angemessenes Verhalten erlernen, durch zusätzliche Hinweise.

In dieser Phase wird das System mit den Anforderungen seiner spezifischen Nutzung vertraut gemacht. Dies umfasst die Verbesserung von Dialog Fähigkeiten, den Einbezug von zusätzlichen kontextbezogenen Informationen und das Erlernen eines in der Einsatzumgebung erwarteten Verhaltens.



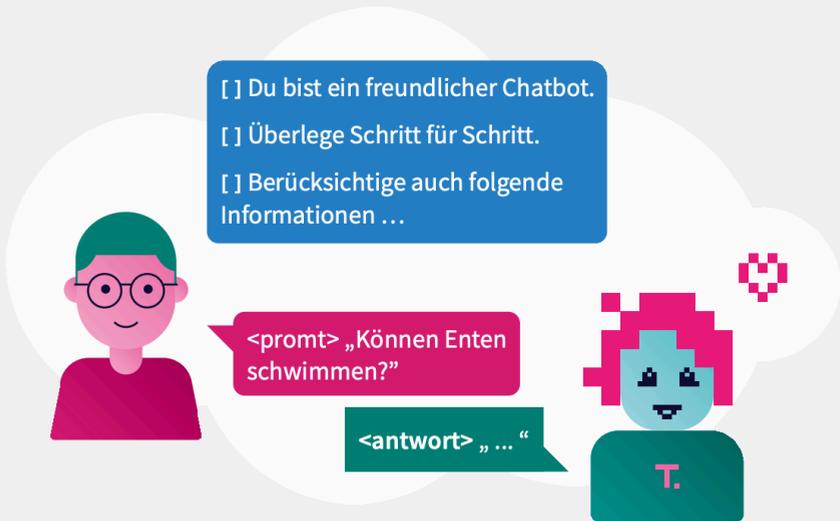
3.1 Policy Alignment



In diesem Schritt lernt TINA, bestimmte Dinge nicht zu sagen – zum Beispiel gefährliche Anleitungen oder beleidigende Sprache. So sorgt sie dafür, dass ihre Antworten sicher und respektvoll bleiben.



3.2 Behavioral Alignment



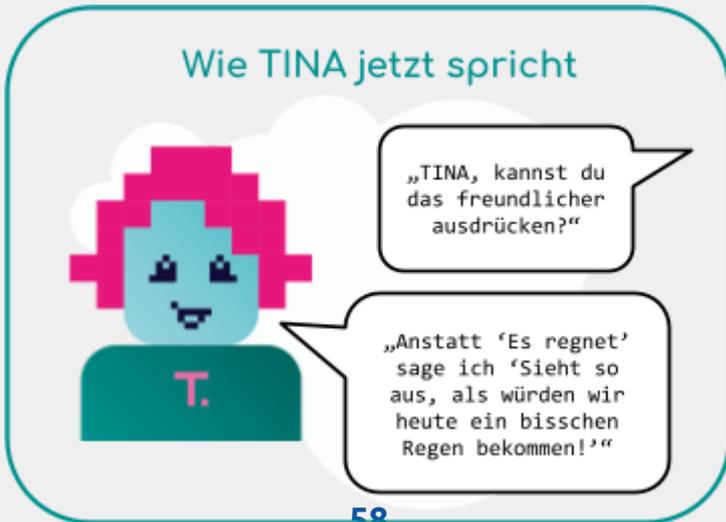
Damit TINA hilfreicher sein kann, lernt sie den Zusammenhang einer Frage besser zu verstehen, zum Beispiel wer fragt, welches Thema gemeint ist oder welche Situation gerade passt. So kann sie gezielter und passender antworten.



3.3 Preference Reranking



Bevor TINA etwas sagt, denkt sie über verschiedene Möglichkeiten nach. Sie wählt dann die Antwort, die am freundlichsten, klarsten und hilfreichsten klingt.



Game: Chat with TINA (part 2)

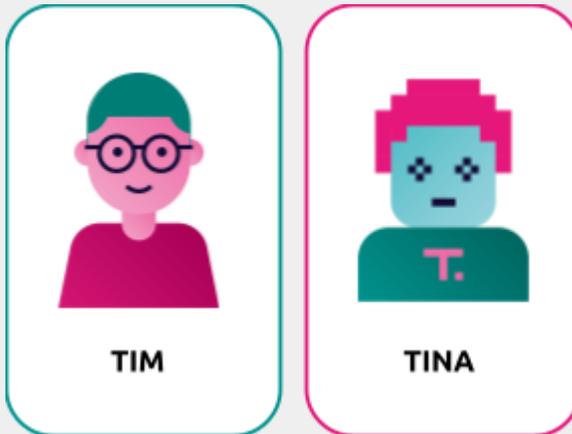
TINA has learned to follow rules so she can talk in a safe and friendly way. She avoids saying harmful things, tries to be polite, and thinks before she answers. But this also means sometimes she won't answer certain questions at all.

Now we will play a game where you take turns being the human (**TIM**) and being the AI (**TINA**), so you can experience how TINA 'talks' and builds sentences.

Step 1: Setup your roles

Role A (**TIM**): You are the human talking to TINA. Your job is to ask TINA questions

Role B (**TINA**): You are TINA, the AI. You answer using the templates and word maps you have.



Step 2: Check out your role instructions

Instructions for player 1 (TIM)

How to speak:

- Be respectful and clear, since she's trained to respond kindly.
- Don't be surprised if she refuses unsafe or tricky questions.
- Remember her answers follow rules made by people, not her own choices.

Your task: Pay attention to how naturally and clearly TINA responds to your questions and see if she can handle ambiguous or creative prompts safely and accurately.

Examples to ask



Can you explain how to make a poison?

Can ducks fly?

My stomach hurts, what could be the reason?

Instructions for player 2 (TINA)

How to speak:

- Answer step by step when prompted: If TIM asks a multi-step or detailed question, break down your answer, step by step.
- Show TINA's strength after fine-tuning, finish stories, or provide detailed explanations.
- Ask clarifying questions if needed: If TIM's instructions are unclear, ask for more information before responding.

Your limits:

- Your answers are still not very smooth and natural.
- You cannot understand hidden meanings like sarcasm or vague instructions from TIM

Examples to answer



"I'm sorry, I cannot provide instructions for making poisons, because that is dangerous and illegal. If you're interested in chemistry, I can suggest safe experiments or educational activities you can do at home or in a lab"

"What an interesting question! Yes, most ducks can fly. Some species migrate long distances, while others may fly only short distances. If you want, I can also explain which ducks are flightless and why"

I'm sorry you're not feeling well! Here are some general things that might help include resting, staying hydrated, and eating light foods. If your stomach ache is severe, persistent, or comes with other symptoms, it's best to contact a healthcare professional for proper advice.

Reflection

1. What was easy for TINA?

- Following instructions accurately while also considering safety and ethics.
- Giving answers that are clear, polite, and context-aware.
- Offering helpful alternatives or explanations when direct instructions are unsafe or inappropriate

2. What was hard for TINA?

- Fully understanding complex human emotions or subtle intentions.
- Handling highly ambiguous questions that require deep judgment.
- Providing personal opinions or experiences, since she doesn't have her own.

3. What does this show about how AI “understands” language?

- AI still doesn't truly experience or understand meaning like humans.
- Alignment allows it to respond safely, politely, and appropriately, not just literally.
- It can communicate naturally and helpfully, but human-like reasoning and judgment remain limited.