



Technische Informationen zum CORA AI-Chat und den verwendeten Sprachmodellen (LLMs)

Die Modelle von OpenAI und den anderen Herstellern von AI-Chats und Wissensmodellen, sind öffentlich besprochen. Alle AI-Chats, welche direkt als Lösung in einer App oder über den Desktop angeboten werden, egal ob gratis oder als professionelle Bezahlversion, sind voreigestellt und lassen nur eine beschränkte Einflussnahme durch die Anwenderperson zu.

Nur Lösungen, welche die Sprachmodelle über eine API (Schnittstelle) ansprechen, können das Verhalten der Sprachmodelle signifikant beeinflussen. Diese Steuerung der LLMs ist für das optimale Funktionieren in einer professionellen Umgebung, wie in einer Firma, absolut notwendig.

Noch wichtiger als das optimale LLM ist die wichtige Frage

Wie dynamisch und kreativ darf eine AI sein, - wenn aufgrund der Antworten aus der AI-Entscheidungen gefällt und Prozesse gesteuert werden sollen? Wie kann man die AI-Systeme steuern und kontrollieren, ohne sie gleichzeitig einzuschränken?

Kleine AI-Regler mit hohem Einfluss

Egal ob interne oder externe Daten und Wissensbestände genutzt werden – sie müssen auf die für die Beantwortung relevanten Stellen verdichtet werden. Zusätzlich spielen in der Welt der AI-Textgeneratoren zwei Regler eine große Rolle: *Temperature* und *Top-p*.

- Die *Temperature* bestimmt, wie „mutig“ die AI bei der Wortwahl ist – bei niedrigen Werten bleibt sie brav beim wahrscheinlichsten nächsten Wort, bei hohen Werten wagt sie kreative Sprünge.
- *Top-p* dagegen legt fest, wie groß der „Wortpool“ ist, aus dem die AI überhaupt auswählt: Bei 1.0 steht das gesamte Vokabular offen, bei kleinen Werten nur die engsten Favoriten.

Beide Parameter beeinflussen sehr stark, ob eine Antwort nüchtern und vorhersehbar oder bunt und überraschend ausfällt.

- Hohe *Temperature* mit hohem *Top-p* gleicht einer Improvisationsbühne, von poetisch bis kurios.
- Niedrige Werte bei beiden führen zu fast identischen Antworten, selbst wenn die gleiche Frage mehrmals gestellt wird.

Faszinierend ist, dass schon kleine Veränderungen große Wirkung haben: Eine kleine Änderung der «*Temperature*»- Einstellung kann den Ton einer Antwort komplett verändern. So zeigt sich, dass Kreativität bei AI nicht nur eine Frage des Trainings ist – sondern auch der Feinjustierung dieser beiden unscheinbaren Regler.

Zugängliche AI-Regler

In CORA (*Produktiv-Version*) kann die Nutzerperson steuern, ob die Antworten eher breit gestreut & kreativ oder fokussiert & vorhersehbar sind. Es kann zwischen 3 Einstellungen gewählt werden, welche die AI-Regler verändern.

Strikt	Exakte Übereinstimmung mit Frage, enges Vokabular, wiederholbare Antworten, wenig Varianz
Moderat	breites Vokabular, Überraschungen, Balance aus Kreativität und Konsistenz,
Kreativ	Maximale Kreativität, unerwartete, ev. chaotische, manchmal unlogische, Antworten, muss manuell überprüft werden!