

# Sachbericht zum Vorhaben 16 IS 22068 – EQUIPE

## Teil I – Kurzbericht

Moderne Verfahren der Künstlichen Intelligenz, insbesondere sogenannte Transformer-Modelle, erzielen heute herausragende Ergebnisse bei der Verarbeitung von Sprache, Bildern und Zeitreihen. In vielen praktischen Anwendungen wie z.B. der Vorhersage des Strombedarfs, reicht jedoch eine reine Punktvorhersage nicht aus. Entscheidend ist zusätzlich die Frage, wie sicher oder unsicher eine Vorhersage ist. Die systematische Quantifizierung solcher Unsicherheiten (engl. Uncertainty Quantification; UQ) ist eine zentrale Voraussetzung für vertrauenswürdige KI-Systeme.

Zum Zeitpunkt des Projektstarts existierten bereits verschiedene Ansätze zur Unsicherheitsquantifizierung, darunter Ensemble-Methoden wie Monte Carlo Dropout sowie Bayes'sche neuronale Netze. Diese Verfahren waren jedoch entweder nur eingeschränkt auf moderne Transformer-Architekturen übertragbar oder mit erheblichem Rechenaufwand verbunden. Insbesondere fehlten skalierbare Konzepte, um Unsicherheitsmethoden effizient auf größere Modelle und realistische Datensätze anzuwenden. Das Projekt EQUIPE setzte genau hier an. Ziel war es, Methoden zur Quantifizierung von Unsicherheiten in Transformer-Netzwerken systematisch zu untersuchen, weiterzuentwickeln und hinsichtlich ihrer Skalierbarkeit und Effizienz zu analysieren. Ein besonderer Fokus lag auf Zeitreihenvorhersagen, beispielsweise im Kontext der Energieforschung.

Biele Methoden zur Unsicherheitsquantifizierung beruhen auf statistischen Sampling-Verfahren, bei denen wiederholt leicht variierte Modellversionen ausgewertet werden, um eine Wahrscheinlichkeitsverteilung der Vorhersage zu approximieren. Diese Verfahren sind jedoch rechnerisch aufwendig, da für eine einzige Vorhersage häufig zahlreiche Modell-Durchläufe erforderlich sind. Zudem war unklar, wie viele solcher Wiederholungen tatsächlich notwendig sind, um eine zuverlässige Unsicherheitsabschätzung zu erhalten. Auch Fragen der Kalibrierung, welche beschreibt ob die vorhergesagte Unsicherheit tatsächlich mit der realen Fehlerwahrscheinlichkeit übereinstimmt, waren für Transformer-Modelle noch nicht systematisch untersucht.

Zu Beginn des Projekts wurde ein eigenes Transformer-Modell für Zeitreihenvorhersagen entwickelt, der sogenannte Residual Cyclic Transformer (ReCycle). Dieses Modell kombinierte eine hohe Vorhersagegenauigkeit mit deutlich reduziertem Rechenaufwand. Diese Entwicklung schuf eine effiziente Grundlage für alle weiteren Arbeiten. Parallel dazu wurde der Stand der Forschung zu Unsicherheitsmethoden umfassend analysiert. Untersucht wurden unter anderem Monte Carlo Dropout, Bayes'sche neuronale Netze mit Variational Inference sowie neuere probabilistische Modellierungsansätze. Dabei zeigte sich, dass viele Methoden grundsätzlich sowohl für Klassifikation als auch für Regression geeignet sind, sofern sie mathematisch korrekt formuliert werden.

Im weiteren Verlauf wurde ein zentrales technisches Problem identifiziert: Der größte Rechenaufwand bei probabilistischen neuronalen Netzen entsteht durch das wiederholte Ziehen von Zufallsstichproben während Training und Inferenz. Um dieses Bottleneck besser zu verstehen, wur-

de erstmals systematisch untersucht, wie viele solcher Stichproben tatsächlich notwendig sind. Dabei konnte ein grundlegender Unterschied zwischen Klassifikations- und Regressionsaufgaben gezeigt werden. Für Klassifikation genügt in der Regel eine einzelne Stichprobe, während für Regression mehrere Durchläufe erforderlich sind. Diese Erkenntnis ermöglicht eine deutliche Reduktion des Rechenaufwands ohne Einbußen bei der Qualität der Unsicherheitsabschätzung.

Darüber hinaus wurde ein Verfahren für paralleles Sampling entwickelt, das es erlaubt, mehrere Stichproben gleichzeitig auf unterschiedlichen Recheneinheiten zu berechnen. Dieses Verfahren verbessert die Skalierbarkeit probabilistischer Modelle erheblich und macht deren Einsatz auch in größeren Anwendungen praktikabel. Ein weiteres zentrales Ergebnis des Projekts war die Entwicklung des Open-Source-Frameworks torch-blue. Dieses Softwarewerkzeug ermöglicht es, bestehende neuronale Netze automatisch in Bayes'sche Modelle zu überführen, ohne dass Anwenderinnen und Anwender tief in die mathematischen Details der Variational Inference einsteigen müssen. Die Schnittstelle orientiert sich an der weit verbreiteten PyTorch-Umgebung, sodass eine nahtlose Integration in bestehende Projekte möglich ist.

Zusätzlich wurden Fragen der Modellkalibrierung untersucht. Außerdem wurde analysiert, inwieweit sogenannte „dünnbesetzte“ Netzwerke auch im bayesschen Kontext möglich sind. Es konnte gezeigt werden, dass auch probabilistische Modelle strukturelle Redundanzen enthalten, die theoretisch reduziert werden können. Allerdings stellte sich heraus, dass aktuelle Software- und Hardware-Infrastrukturen echte Rechenvorteile durch Sparsity nur eingeschränkt unterstützen.

Das Projekt EQUIPE hat wesentliche Fortschritte bei der effizienten und skalierbaren Quantifizierung von Unsicherheiten in Transformer-Modellen erzielt. Zu den zentralen Ergebnissen zählen die Entwicklung eines besonders effizienten Transformer-Modells für Zeitreihen, die Identifikation und Reduktion des Sampling-Bottlenecks, die Entwicklung eines skalierbaren Sampling-Parallelismus sowie die Bereitstellung des Software-Frameworks torch-blue als Open-Source-Lösung. Die Ergebnisse wurden auf internationalen Konferenzen präsentiert und als Open-Source-Software veröffentlicht.

Durch diese Beiträge wurde die praktische Anwendbarkeit probabilistischer Deep-Learning-Methoden deutlich verbessert. Dies ist insbesondere für Anwendungen relevant, in denen verlässliche Unsicherheitsangaben erforderlich sind, etwa in der Energie- und Klimaforschung oder bei industriellen Prognosesystemen. Darüber hinaus hat das Projekt zur strukturellen Stärkung der KI-Forschung beigetragen, indem eine eigenständige Nachwuchsgruppe aufgebaut und langfristig etabliert wurde. Insgesamt leistet EQUIPE damit einen wichtigen Beitrag zur Entwicklung vertrauenswürdiger, effizienter und skalierbarer KI-Systeme, die nicht nur Vorhersagen liefern, sondern auch deren Unsicherheit transparent machen.