

Berufsakademie Sachsen
Staatliche Studienakademie Dresden
Hans-Grundig-Straße 25
01307 Dresden



Bericht zum
Workshop Machine Learning und Deep Learning
Zeit: 6. November, 2024, 9:45 – 16:10 Uhr

Dresden, 19.11.2024

Am 6. November fand an der Staatlichen Studienakademie ein Workshop zu künstlicher Intelligenz statt. Thematisiert wurden aktuelle Entwicklungen und Anwendungen im Bereich Machine Learning und Deep Learning, der mit den diesjährigen Nobelpreisen für Physik und Chemie geehrt wird. Die öffentliche, vom Förderverein der Staatlichen Studienakademie Dresden e.V. organisierte Veranstaltung richtete sich an alle Studierenden, Dozierende und Praxispartner sowie weitere Personen, die daran interessiert waren, ein tieferes Verständnis für die neuesten Technologien und Trends im Bereich Künstlicher Intelligenz (KI) zu gewinnen. An der hybrid durchgeführten Veranstaltung, nahmen rund 100 Personen teil, etwa 40 online.

Eröffnet wurde die Veranstaltung mit einem Grußwort des Vorstandsvorsitzenden des Fördervereins, Herrn Dr. Kleineidam. Es folgte ein **Einführungsvortrag** von Prof. Dr. Daniel Gembris mit einem Überblick zum Workshop-Programm, einer Erläuterung des Prinzips künstlicher neuronaler Netze anhand des Hopfield-Netzes und eine Gegenüberstellung des Nutzens für die Nachhaltigkeit und den steigenden Energiebedarf. Nachhaltigkeit von KI war Schwerpunktthema von vier der sieben nachfolgenden Vorträge und damit auch das Leitmotiv des gesamten Workshops.

In seinem Vortrag „**Generative KI: Von Transformer Architekturen zu Diffusion Modellen. Grundlagen und Anwendungen.**“ gab Professor Dr. Frank Schönefeld von unserem Praxispartner T-Systems Multimedia Solutions einen Überblick über die aktuell für Deep Learning eingesetzten Algorithmen und Netzwerk-Architekturen. Zunächst wurde der interdisziplinäre Charakter von Machine Learning und Deep Learning und deren wichtigste Beschreibungsdimensionen aufgezeigt: Charakteristiken künstlicher Neurone, Netz-Topologie, Lernmethode und Lernregeln. Anhand von Beispielen wurden zwischen Diskriminativer und Generativer KI sowie zwischen Fluss- und Energiebasierten Modellen unterschieden. Es folgte eine genauere Vorstellung der Transformer-Architekturen, die die Grundlage für die zurzeit sehr populären Large Language Models, die sich z.B. hinter ChatGPT oder google gemini verbergen, und der Diffusionsmodelle, die von vielen KI-Bildgeneratoren, wie z.B. DALL-E, aber auch FLUX.1 des deutschen KI-Startups Black Forrest Labs, genutzt werden. Trotz der zunehmenden Fähigkeiten von KIs Programmcode zu erstellen, als Beispiel wurde google Alphacode 2 genannt, würden menschliche Entwickler weiterhin benötigt. Die KI ersetze diese nicht, sondern unterstütze sie. Mit zunehmender Komplexität der Aufgaben sinke aber der mit KI erzielbare Zeitgewinn.

Prof. Dr. Steffen Löck vom Strahlentherapiezentrum OncoRay in Dresden fokussierte in seinem Vortrag „**KI in der Medizin**“ auf Anwendungen in der Strahlentherapie, bei der Tumorgewebe durch konzentrierte Teilchenstrahlen zerstört wird. Protonen erweisen sich für eine Strahlentherapie als besonders vorteilhaft. Über die entsprechende technische Ausstattung verfügen aber nur wenige medizinische Einrichtungen in Deutschland: Diese befinden sich neben Dresden in Essen, Heidelberg, München und Berlin. Ziel seiner Arbeitsgruppe ist es, den Behandlungserfolg der Strahlentherapie, der sich sowohl auf die Tumorheilung als auch auf die Nebenwirkungen bezieht, mittels statistischer Modellierung vorherzusagen. Dies soll die Grundlage bieten, um Patienten noch gezielter für individuelle Therapien auswählen zu können. Zur Erstellung der Modelle werden klinische Daten sowie genetische und radiologische Informationen verwendet. Da für jeden Patienten viele Variablen zur Verfügung stehen, werden maschinelle Lernverfahren eingesetzt, um (a) die wichtigsten Variablen zu finden (Signatur) und (b) Modelle zur Vorhersage des klinischen Erfolgs aufzustellen. Es wurden Ergebnisse für ein Modell präsentiert, das Daten von circa 300 Patienten, u.a. zur Bildgebung des Tumors in der Computertomographie, zu Alter, Geschlecht und Tumorstadium, mit Informationen zum Wiederauftreten von Kopf-Hals-Tumoren verknüpft. Der Einsatz von Deep Learning konnte die Prognose-Qualität gegenüber konventionellen Machine-Learning-Ansätzen moderat verbessern. Aktuell erfolgt der Aufbau des MEDomics-Konsortiums, in dem internationale Partner anonymisierte Daten austauschen, um die Basis an KI-Trainingsdaten zu vergrößern (siehe: https://youtu.be/2030Pdgm3_4). In einer ersten Veröffentlichung des Konsortiums konnte gezeigt werden, dass Deep-Learning-basierte Spracherkennung von ärztlichen Befunden die Vorhersage des Überlebens nach der Behandlung von Brustkrebs signifikant verbessern konnte.

Es folgte ein Vortrag von Prof. Dr. Michael Färber vom ScaDS.AI, dem Center for Scalable Data Analytics and Artificial Intelligence, mit den beiden Standorten Dresden und Leipzig zu „**KI x Nachhaltigkeit: Modellierung des Energieverbrauchs von KI-Modellen und Vorhersage von ESG-Bewertungen für Unternehmen**“. Seine Forschung ist in der Schnittmenge von Wissensrepräsentation, Maschinen-Lernen und der Verarbeitung natürlicher Sprache angesiedelt. Die seinem Vortrag vorangestellte Leitfrage lautete: „Wie können wir Computer trainieren, damit sie große Textmengen lesen und verstehen, die Informationen effektiv strukturieren und diese für die Entwicklung von transparenteren und verlässlicheren Empfehlungssystemen nutzen?“. Eine wichtige Motivation für seine Arbeit ist die rasant zunehmende Zahl von wissenschaftlichen Veröffentlichungen, die bereits Anfang 2010er Jahre die Zahl von 10 Millionen pro Jahr überschritten hat, also selbst in einzelnen Themengebieten viel mehr als einzelne Menschen lesen können. Deshalb wurden Systeme für die Empfehlung von Publikationen, von Quellenangaben, Datensätzen und die Beantwortung wissenschaftlicher Fragen entwickelt. Mit SimplifyMyText steht ein Werkzeug für Übersetzung von Webseiten in einfache Sprache zur Verfügung (<https://SimplifyMyText.org>). Für diese Projekte werden sog. Wissensgraphen genutzt, die Beziehungen zwischen Datenbankeinträgen beschreiben. Als Beispiel wurde eine Person genannt, die sich für ein bestimmtes Kunstwerk interessiert, das von einer anderen Person geschaffen wurde. Weit verbreitete Wissensgraphen seien DPpedia, Wikidata, Yago und product graph. Das weltweit größte System sei SemOpenAlex (<https://semopenalex.org/>) mit mehr als 26 Milliarden RDF(Resource Description Framework)-Einträgen, die aus Subjekt, Prädikat und Objekt bestehen. Woran es bislang fehlte, war die Extraktion wissenschaftlicher Schlüsselinformationen, z.B. zu den verwendeten Datensätzen, Metrik, Aufgabe und Methode. Mit LPWC (Linked Papers With Code) steht jetzt aber ein entsprechender RDF-Graph zu mehr als 400.000 Maschinenlernen-Veröffentlichungen zur Verfügung, die sich automatisiert konstruieren lassen.

Der zweite Teil des Vortrags thematisierte dann den Energiebedarf von KI. ChatGPT alleine benötige aktuell jährlich 453,6 Millionen kWh elektrische Energie für die Reaktion auf Prompts, die z.B. ausreichte um ganz Österreich für 2,5 Tage mit Strom zu versorgen. Die Verfügbarkeit von Grafikkarten mit einer für KI-Berechnungen prädestinierten Hardware-Architektur hat seit den 2010er Jahren zu einer schnelleren Zunahme der ausgeführten KI-Rechenoperationen und gleichzeitig zu einem schnelleren Anstieg der Energienachfrage der Rechenzentren geführt. Die aller Rechenzentren weltweit übersteige bereits die vieler Länder. Für die Abschätzung des Rechenaufwands und Energiebedarfs von KI wurde die Green AI Ontologie entwickelt. Unter einer Ontologie wird in der Informatik eine meist sprachlich formulierte und formal geordnete Darstellung einer Menge von Begriffen und der zwischen ihnen bestehenden Beziehungen verstanden. Mit der entwickelten Ontologie werden verschiedene Aspekte modelliert: Metriken und Werkzeuge, verwendete Hardware, installierte Software und Literaturdaten. Ein weiteres Projekt der Arbeitsgruppe von Professor Färber hat zum Ziel, die ESG-Nachhaltigkeitsbewertung von Unternehmen aus öffentlich zugänglichen Informationen abzuleiten. Die einzelnen Buchstaben in „ESG“ stehen für environmental, social und governance. Gemeint sind damit Auswirkungen auf die Umwelt, soziale Verantwortung und eine verantwortungsvolle Unternehmensführung.

Professor Loviscach von der Hochschule Bielefeld, der in den Medien als YouTube-Professor bekannt ist, ließ mit seinem eingängigen Vortrag „**Prompten und prompten lassen: KI in der Hochschullehre**“ nach der Mittagspause erst gar keine Mittagsmüdigkeit aufkommen. Seine Agenda umfasste fünf Punkte: Automatisierung, wie z.B. die automatisierte Erstellung von Klausuren und Musterlösungen oder die Betreuung durch KI, neue Möglichkeiten, u.a. durch LLMs mit Spracherkennung und der Vorgaben zum Antwort-Stil, die Leistungsentwicklung von KIs beim Lösen von Klausuraufgaben, der Stärkung von Privilegien durch KI und die Zukunft verschiedener Berufe vor dem Hintergrund immer leistungsstärkerer KIs. Eine neue, im Vortrag diskutierte Art von Täuschungsversuch ist die Benutzung funkvernetzter In-Ohr-Hörgeräte, mit denen sich Studierende Lösungen von Klausuraufgaben vorsagen lassen könnten. Diese Aussicht kulminierte in der etwas provokativen Frage, ob demnächst vor Klausuren die Gehörgänge von Studierenden untersucht werden müssten. In der Berufswelt sei damit zu rechnen, dass viele Routineaufgaben wegrationalisiert werden. Weil humanoide Roboter noch teuer und nicht sehr alltagstauglich sind, hätten Handwerker zumindest für einige Zeit noch eine Schonfrist. Das von einigen gehypte Berufsbild des Prompt-Engineerings wird von Professor Loviscach kritisch bewertet.

In seinem Vortrag „**Energieeinsparung in der KI mittels Neuromorphic Computing**“ berichtete Bernhard Vogginger von der TU Dresden über neue Hardware für energieeffizientere KI. Ein Schwerpunkt lag dabei auf der neuen "Spiking Neural Network Architecture", kurz SpiNNaker, welche auf die Simulation neuronaler Netze optimiert ist. Diese integriert eine große Zahl von ARM-Prozessoren, die auch in Smartphones und Tablets eingesetzt werden, und ist für Kommunikation und Speicherzugriff optimiert. Dies ermögliche energieeffiziente Simulation in biologischer Echtzeit, wobei SpiNNaker bezüglich beider Aspekte aktuellen Großrechnern deutlich überlegen sei. Die erste Generation von SpiNNaker wurde von der University of Manchester entwickelt und bis zu einer Systemgröße von 1 Millionen ARM Prozessoren skaliert. Dies erlaube die Echtzeitsimulation (1 ms Latenzzeit) von Netzen in der Größenordnung von 1 % des menschlichen Gehirns. Nach der Darlegung der Motivation und einer Erläuterung von Neuromorphic Computing wurde direkt das aktuelle Projekt SpiNNcloud vorgestellt, in dem eine SpiNNaker2-Maschine mit mehr als 5 Millionen ARM Prozessoren an der TU Dresden aufgebaut wird.

Neben der Nutzung zur Gehirnsimulation, soll diese in Sachsen als ‚SpiNNcloud‘ für Forschung in den Bereichen energieeffiziente KI, Echtzeit-KI (z.B. für taktiles Internet, Industrie 4.0 und automotive) und der computergestützten Wirkstoffentwicklung etabliert werden. Die Ansiedlung von Unternehmen der Chip-Industrie in Dresden und Umgebung hat dazu geführt, dass von „Silicon Saxony“ gesprochen wird. Die SpiNNaker2-Entwicklung hat in großem Maße vom Technologie-Ökosystem in Sachsen profitiert, unter anderem durch das Unternehmen Cloud&Heat, das mit einem eigenen Vortrag auf dem Workshop vertreten war. Das ebenfalls in Dresden angesiedelte Spin-off SpiNNcloud Systems GmbH übernimmt die Kommerzialisierung und Weiterentwicklung der SpiNNaker2-Hardware und unterstreicht die Wettbewerbsfähigkeit Sachsen bei der Entwicklung von KI-Hardware.

Der Vortrag von Frau Dr. Dr. Friederike Rohde vom Berlin Ethics Lab der TU Berlin und dem Institut für ökologische Wirtschaftsforschung (IÖW) mit dem Titel „**Maschinelles Lernen aus Nachhaltigkeitsperspektive**“ betrachtete KI hinsichtlich der verschiedenen Nachhaltigkeits-Dimensionen: die ökologische, die soziale und die ökonomische. Im Vortrag wurden dazu empirische Daten zur wachsenden Größe von Large Language Modellen, dem Stromverbrauch und dem Ressourcenaufwand für die Hardware-Produktion präsentiert. Viele Teilnehmende waren sicherlich überrascht, dass KI mit einem nicht unerheblichen Wasserverbrauch KI verbunden ist: Große Rechenzentren werden nämlich für eine bessere Kühlleistung oft nicht mit Luft, sondern durch die Verdunstung von Wasser gekühlt. Das Wasser ist dadurch nicht verloren, weil der Wasserdampf später wieder kondensiert und als Regen auf die Erde fällt - dennoch verschärft sich die Konkurrenz um Wasser: Das für die Kühlung genutzte Wasser fehlt, wie im Vortrag gezeigt, z.B. in der Landwirtschaft. Für die Reduktion des Energieverbrauchs wurden verschiedene Lösungsansätze genannt, darunter der Einsatz von kleineren Netzen, die mittels größerer Netze trainiert werden. Im Vortrag wurde auf verschiedene Tools für die Bestimmung der Treibhausgas-Emissionen von KI hingewiesen. Das Thema lokale gehostete KI-Software wurde im letzten Vortrag aufgegriffen und vertieft.

Der letzte Vortragende war Patrick Thiem vom BA-Praxispartner Cloud&Heat, einem Dresdner Unternehmen für Cloud-Computing. Eines der Geschäftsmodelle von Cloud&Heat besteht in der Entwicklung, Vertrieb und Installation von Technik für die Nutzung der Abwärme von Rechenzentren für das Heizen und die Bereitstellung von Warmwasser. Weiterhin befasst sich Cloud&Heat auch mit der Konzeption, Entwicklung und dem Betrieb leistungsfähiger Open Source Systeme. In seinem Vortrag mit dem Titel "**Kein Mutterschiff, kein Problem: Datenanalyse mit lokaler KI stärken**" wurde von einer Proof-of-Concept-Lösung für die Bearbeitung von Supporttickets berichtet, für die eine lokal gehostete KI-Software („On-Premise“) verwendet wurde. Die lokale Ausführung von KI-Systemen kann sich besonders wegen der strengen Datenschutzbestimmungen in Deutschland als entscheidender Wettbewerbsvorteil hiesiger Unternehmen erweisen. Die Unabhängigkeit von großen internationalen Cloud-Anbietern eröffnete zudem neue Möglichkeiten der Systemanpassung – das Modell könne gezielt auf spezifische Geschäftsprozesse optimiert und nahtlos in die bestehende IT-Landschaft integriert werden. Der Vortrag ging auch auf die praktische Umsetzung mittels eines vollständig Open-Source-basierten Ansatzes ein, bei dem das Large Language Model Llama3.1 70b zum Einsatz kam. In Kombination mit Retrieval Augmented Generation (RAG) – also das KI-basierte Generieren von Text, ergänzt durch das Abrufen von Informationen und der Nutzung historischer Supportticket-Daten, sei ein leistungsfähiges System entstanden, das auf zwei Nvidia A10 GPUs betrieben wurde.

Diese Hardware-Konfiguration ermögliche eine vollständige Kontrolle über die Infrastruktur und gewährleiste gleichzeitig die nötige Skalierbarkeit für wachsende Unternehmensanforderungen. Das Fazit: Die Kombination aus Datenschutz, Anpassungsfähigkeit und Nachhaltigkeit mache lokale KI zu einem zukunftsweisenden Modell für die digitale Transformation.

Weiterführende Informationen: Siehe BA-Website

Prof. Dr. Daniel Gembris, Dresden, 19.11.2024

Förderverein der Staatlichen Studienakademie Dresden e.V.

Der Förderverein wurde 2015 gegründet. Selbstloser Zweck des Vereins ist die Förderung akademischer, sozialer, kultureller und sportlicher Interessen der Studierenden.

Deshalb unterstützt der Förderverein die Durchführung wissenschaftlicher Veranstaltungen und Forschungsvorhaben, die Gemeinschaft der Studierenden, die Vorbereitung von Feierlichkeiten zum erfolgreichen Studienabschluss sowie natürlich den Werdegang der Studierenden.



Die Mitgliedsbeiträge und Spenden sowie die ehrenamtliche Tätigkeit der Mitglieder unterstützen die Realisierung dieser Vorhaben. Der Förderverein ist daher sehr auf finanzielle Unterstützung angewiesen.

Wenn auch Sie Interesse daran haben, mit Ihrem Beitrag und gegebenenfalls auch durch aktive Beteiligung die Studierenden der Staatliche Studienakademie Dresden zu unterstützen, dann freuen wir uns auf Ihren Antrag auf Mitgliedschaft im Förderverein!

Der jährliche Mitgliedsbeitrag ist für natürliche Personen ab 30 €, für juristische Personen ab 200 € frei gestaltbar. Als gemeinnütziger Verein stellen wir gerne Zuwendungsbescheinigungen aus.

Weitere Informationen zum Verein unter: www.ba-dresden.de/foerdereverein

Spenden gerne direkt auf unser Bankkonto bei der ostsächsischen Sparkasse Dresden (siehe Fußzeile).