

# **Philosophie der AI**

Gerd Graßhoff

2024-05-18

# Table of contents

<b>Vorlesung <i>Philosophie der AI</i></b>	<b>13</b>
<b>1 Was ist AI?</b>	<b>14</b>
<b>2 Begrüßung und Einführung</b>	<b>15</b>
2.1 Was ist AI? . . . . .	16
<b>3 AI als Alleskönner</b>	<b>17</b>
3.1 Der Durchbruch der AI-Visionen . . . . .	17
3.2 Die Attraktivität von AI . . . . .	18
3.3 Die ursprüngliche Idee des Internets . . . . .	18
3.4 Die Ablösung der Webwelt durch AI . . . . .	18
3.5 Die Umgestaltung der Architektur des Internets . . . . .	19
<b>4 Neue Möglichkeiten durch Künstliche Intelligenz</b>	<b>20</b>
4.1 Hochwertige Übersetzungen . . . . .	20
4.2 Simultanübersetzung und Lektoratsassistenz . . . . .	20
4.3 Automatisierte Forschungsberichte . . . . .	21
4.4 Das Labor für Lettre AI . . . . .	21
4.5 Der Kern der Vorlesung . . . . .	23
4.6 Übertragen eines Bildes in maschinenlesbaren Text . . . . .	23
4.7 Übersetzen des Textes in eine andere Sprache . . . . .	23
<b>5 Erweiterung der Möglichkeiten durch Phantasie und gezielte Fragestellungen</b>	<b>24</b>
5.1 Analogie zu Sherlock Holmes . . . . .	24
5.2 Vielfältige Analysemöglichkeiten von Texten . . . . .	24
<b>6 Philosophie als Grundlage für die Möglichkeiten der AI</b>	<b>25</b>
6.1 Beantwortung von Fragen über Mikrofoneingabe . . . . .	25
6.2 Die Möglichkeiten der AI . . . . .	25
6.3 Gefahren der AI . . . . .	26
6.4 Der sprachliche Kern der AI . . . . .	26
6.5 Das Problem der Halluzinationen . . . . .	27
6.6 Die Gefahr der Manipulation durch glaubwürdige Fakes . . . . .	27
6.7 Selektive Informationen und die Pluralität der Hintergründe . . . . .	28
6.8 Die Unausweichlichkeit der AI-Entwicklung und die Notwendigkeit der Gestaltung	28

6.9 Weitere Gefahren: Diskriminierung und Überwachung . . . . .	28
6.10 Die Notwendigkeit der Auseinandersetzung mit AI . . . . .	29
<b>7 Nutzungsmöglichkeiten in der Wissenschaft</b>	<b>30</b>
<b>8 Bislang nicht lösbarer Aufgaben</b>	<b>31</b>
8.1 Frage 1: Einfache Aussage in einer Quelle . . . . .	31
8.2 Frage 2: Aussage in Briefen zu einem Thema . . . . .	31
8.3 Frage 3: Aussagen einer Person in ihren Schriften . . . . .	32
8.4 Frage 4: Keine Aussage einer Person in ihren Schriften . . . . .	32
<b>9 Die Herausforderung der inhaltlichen Analyse mit AI</b>	<b>33</b>
9.1 Grenzen der traditionellen Datenbanken . . . . .	33
9.2 Qualifizierte Aussagen auf Basis der verfügbaren Evidenz . . . . .	34
9.3 Herausforderungen bei der Interpretation von Metaphern und Ironie . . . . .	34
9.4 Lernfähigkeit und Entwicklungspotenzial von AI-Systemen . . . . .	35
9.5 Der Paradigmenwechsel durch Large Language Models und Embeddings . . . . .	35
9.6 Die Bedeutung der Philosophie für die AI-Forschung . . . . .	35
<b>10 Philosophie der AI</b>	<b>37</b>
<b>11 Begrüßung und Einführung in die Vorlesung "Philosophie der AI"</b>	<b>38</b>
11.1 Die Rolle der Philosophie in der AI . . . . .	38
<b>12 Die KI-Revolution und ihre Auswirkungen</b>	<b>39</b>
12.1 Eine technologisch-gesellschaftliche Revolution . . . . .	39
12.2 Herausforderungen in der Vorlesungsvorbereitung . . . . .	39
<b>13 Vorkenntnisse und Erwartungen an die Studierenden</b>	<b>40</b>
13.1 Vertrautheit mit ChatGPT . . . . .	40
13.2 Begriff der AI oder KI . . . . .	40
<b>14 Organisatorisches und Tools</b>	<b>41</b>
14.1 Vorlesungszeiten und -ort . . . . .	41
14.2 Vorlesungswebseite und Materialien . . . . .	41
14.3 Zugriff auf Chat-GPT . . . . .	41
14.4 Eigene KI-Webseite und virtueller Wittgenstein . . . . .	41
14.5 Moodle und Teilnehmerliste . . . . .	42
14.6 Zulassung und Modulabschlussnoten . . . . .	42
<b>15 Nachfragen und individuelle Anliegen</b>	<b>43</b>
<b>16 Was ist AI?</b>	<b>44</b>
<b>17 KI als Verkaufsargument</b>	<b>45</b>

<b>18 Der Durchbruch der KI-Visionen</b>	<b>46</b>
<b>19 Die Attraktivität von KI</b>	<b>47</b>
19.1 Die ursprüngliche Idee des Internets . . . . .	47
19.2 Die Ablösung der Webwelt durch KI . . . . .	47
19.3 Die Umgestaltung der Architektur des Internets . . . . .	48
19.4 Neue Möglichkeiten durch Künstliche Intelligenz . . . . .	48
19.4.1 Hochwertige Übersetzungen . . . . .	49
19.4.2 Simultanübersetzung und Lektoratsassistenz . . . . .	49
19.4.3 Automatisierte Forschungsberichte . . . . .	49
19.5 Das Labor für gebildete KI . . . . .	50
19.6 Der Kern der Vorlesung . . . . .	50
<b>20 Demonstrieren der Möglichkeiten von ChatGPT anhand eines Beispiels</b>	<b>51</b>
20.1 Übertragen eines Bildes in maschinenlesbaren Text . . . . .	51
20.2 Übersetzen des Textes in eine andere Sprache . . . . .	51
<b>21 Erweiterung der Möglichkeiten durch Phantasie und gezielte Fragestellungen</b>	<b>52</b>
21.1 Analogie zu Sherlock Holmes . . . . .	52
21.2 Vielfältige Analysemöglichkeiten von Texten . . . . .	52
<b>22 Philosophie als Grundlage für die Möglichkeiten der KI</b>	<b>53</b>
22.1 Beantwortung von Fragen über Mikrofoneingabe . . . . .	53
22.2 Die Möglichkeiten der KI . . . . .	53
22.3 Die Gefahren der KI . . . . .	54
22.4 Der sprachliche Kern der KI . . . . .	54
22.5 Das Problem der Halluzinationen . . . . .	55
22.6 Die Gefahr der Manipulation durch glaubwürdige Fakes . . . . .	55
22.7 Selektive Informationen und die Pluralität der Hintergründe . . . . .	56
22.8 Die Unausweichlichkeit der KI-Entwicklung und die Notwendigkeit der Gestaltung	56
22.9 Weitere Gefahren: Diskriminierung und Überwachung . . . . .	56
22.10 Die Notwendigkeit der Auseinandersetzung mit KI . . . . .	57
<b>23 Beispiele für die Nutzung von Sprachen in der Wissenschaft</b>	<b>58</b>
<b>24 Aufgaben und Fragen, die mit herkömmlichen Methoden nicht lösbar sind</b>	<b>59</b>
24.1 Frage 1: Einfache Aussage in einer Quelle . . . . .	59
24.2 Frage 2: Aussage in Briefen zu einem Thema . . . . .	59
24.3 Frage 3: Aussagen einer Person in ihren Schriften . . . . .	60
24.4 Frage 4: Keine Aussage einer Person in ihren Schriften . . . . .	60
<b>25 Die Herausforderung der inhaltlichen Analyse mit KI</b>	<b>61</b>
25.1 Grenzen der traditionellen Datenbanken . . . . .	61
25.2 Qualifizierte Aussagen auf Basis der verfügbaren Evidenz . . . . .	62

25.3 Herausforderungen bei der Interpretation von Metaphern und Ironie . . . . .	62
25.4 Lernfähigkeit und Entwicklungspotenzial von KI-Systemen . . . . .	63
25.5 Der Paradigmenwechsel durch Large Language Models und Embeddings . . . . .	63
25.6 Die Bedeutung der Philosophie für die KI-Forschung . . . . .	63
<b>26 Traditionell schwer lösbare Fragen in der Forschung</b>	<b>65</b>
26.1 Evidenz finden, um eine Hypothese zu widerlegen . . . . .	65
26.2 Zeitgenössische Autoren und ihre Äußerungen zu historischen Hypothesen . . . . .	65
26.3 Der Einfluss von Publikationen auf historische Autoren . . . . .	66
<b>27 Die Rolle der AI in der geisteswissenschaftlichen Forschung</b>	<b>67</b>
27.1 Die Gefahren der AI und ihre Korrektur durch verbesserte Praktiken . . . . .	67
27.2 Nachvollziehbarkeit von Begründungen für historische Hypothesen . . . . .	67
<b>28 Die Entwicklung der KI und ihr Einfluss auf das wissenschaftliche Arbeiten</b>	<b>69</b>
28.1 Die Entwicklung von Interfaces zur Interaktion mit KI . . . . .	69
28.2 Entwicklung der Eingabemöglichkeiten . . . . .	69
28.3 Der Aufstieg von Chat-GPT . . . . .	70
28.4 Weitere Interaktionsmöglichkeiten . . . . .	70
<b>29 Die Architektur hinter den KI-Systemen</b>	<b>71</b>
29.1 Generative KI . . . . .	71
29.2 Von der Syntax zur Semantik . . . . .	71
29.3 Von der Suche nach Zeichenketten zur Suche nach Inhalten . . . . .	72
29.4 Sätze, Aussagen und Wahrheitswerte . . . . .	72
29.5 Die epistemische Dimension des Wissens . . . . .	73
29.6 Die Maschine lernt, Aussagen zu treffen . . . . .	73
29.7 Die Revolution der Sprachmodelle . . . . .	74
29.7.1 Das Prinzip der Embeddings . . . . .	74
29.7.2 Die Bedeutung eines Satzes . . . . .	74
29.8 Die Herausforderung der Bedeutungsgleichheit . . . . .	75
29.8.1 Beispiele für Bedeutungsgleichheit . . . . .	75
29.9 Die Lösung durch künstliche Intelligenz . . . . .	75
29.9.1 Komplexe Übersetzungen . . . . .	75
29.10 Das Training der KI-Modelle . . . . .	76
29.10.1 Die Parameter der Modelle . . . . .	76
29.10.2 Trainingsdatensätze und Übersetzungsliteratur . . . . .	76
29.11 Weitere Trainingsdaten . . . . .	76
29.12 Kontextabhängigkeit der Bedeutung . . . . .	77
29.13 Funktionsweise der KI bei inhaltlichen Fragen . . . . .	77
29.14 Erweiterung der Embeddings auf multimediale Inhalte . . . . .	77

<b>30 Attention is all you need - die zweite Revolution</b>	<b>78</b>
30.1 Transformation von Sequenzen . . . . .	78
30.2 Die Bedeutung der Sequenztransformation . . . . .	78
<b>31 Rettungsversuch und KI-Demonstration</b>	<b>79</b>
31.1 Die Herausforderung der Feststellung . . . . .	79
31.2 Von der Frage zur Anweisung . . . . .	80
31.3 Die Bedeutung der Aufmerksamkeit . . . . .	80
31.4 Die Macht der Kontextualisierung . . . . .	80
31.5 Das Problem der Halluzination . . . . .	81
31.6 Konsequenzen für die Verwendung von KI-generierten Texten . . . . .	81
<b>32 Erweiterung der KI-Modelle um Wissen und Validierung</b>	<b>82</b>
32.1 Notwendige Ergänzungen für sachliche Korrektheit . . . . .	82
32.2 Verhältnis von Sprache und Sachlichkeit . . . . .	82
<b>33 Auswirkungen der KI-Entwicklung auf Sprache und Bedeutung</b>	<b>83</b>
33.1 Zirkularität der Bedeutung in KI-trainierten Texten . . . . .	83
33.2 Übersetzungsfähigkeiten aktueller Programme . . . . .	83
<b>34 Die Bedeutung der Sprachverwendung</b>	<b>84</b>
<b>35 Die Gefahren fehlerhafter Kontexte</b>	<b>85</b>
<b>36 Erfolge und Anwendungen</b>	<b>86</b>
<b>37 Die Bedeutung des Chats</b>	<b>87</b>
<b>38 Philosophie der AI</b>	<b>88</b>
<b>39 Begrüßung und Einführung in die Vorlesung “Philosophie der AI”</b>	<b>89</b>
39.1 Die Rolle der Philosophie in der AI . . . . .	89
<b>40 Die KI-Revolution und ihre Auswirkungen</b>	<b>90</b>
40.1 Eine technologisch-gesellschaftliche Revolution . . . . .	90
40.2 Herausforderungen in der Vorlesungsvorbereitung . . . . .	90
<b>41 Vorkenntnisse und Erwartungen an die Studierenden</b>	<b>91</b>
41.1 Vertrautheit mit ChatGPT . . . . .	91
41.2 Begriff der AI oder KI . . . . .	91
<b>42 Organisatorisches und Tools</b>	<b>92</b>
42.1 Vorlesungszeiten und -ort . . . . .	92
42.2 Vorlesungswebseite und Materialien . . . . .	92
42.3 Zugriff auf Chat-GPT . . . . .	92

42.4 Eigene KI-Webseite und virtueller Wittgenstein . . . . .	92
42.5 Moodle und Teilnehmerliste . . . . .	93
42.6 Zulassung und Modulabschlussnoten . . . . .	93
<b>43 Nachfragen und individuelle Anliegen</b>	<b>94</b>
<b>44 Was ist AI?</b>	<b>95</b>
<b>45 KI als Verkaufsargument</b>	<b>96</b>
<b>46 Der Durchbruch der KI-Visionen</b>	<b>97</b>
<b>47 Die Attraktivität von KI</b>	<b>98</b>
47.1 Die ursprüngliche Idee des Internets . . . . .	98
47.2 Die Ablösung der Webwelt durch KI . . . . .	98
47.3 Die Umgestaltung der Architektur des Internets . . . . .	99
47.4 Neue Möglichkeiten durch Künstliche Intelligenz . . . . .	99
47.4.1 Hochwertige Übersetzungen . . . . .	100
47.4.2 Simultanübersetzung und Lektoratsassistenz . . . . .	100
47.4.3 Automatisierte Forschungsberichte . . . . .	100
47.5 Das Labor für gebildete KI . . . . .	101
47.6 Der Kern der Vorlesung . . . . .	101
<b>48 Demonstrieren der Möglichkeiten von ChatGPT anhand eines Beispiels</b>	<b>102</b>
48.1 Übertragen eines Bildes in maschinenlesbaren Text . . . . .	102
48.2 Übersetzen des Textes in eine andere Sprache . . . . .	102
<b>49 Erweiterung der Möglichkeiten durch Phantasie und gezielte Fragestellungen</b>	<b>103</b>
49.1 Analogie zu Sherlock Holmes . . . . .	103
49.2 Vielfältige Analysemöglichkeiten von Texten . . . . .	103
<b>50 Philosophie als Grundlage für die Möglichkeiten der KI</b>	<b>104</b>
50.1 Beantwortung von Fragen über Mikrofoneingabe . . . . .	104
50.2 Die Möglichkeiten der KI . . . . .	104
50.3 Die Gefahren der KI . . . . .	105
50.4 Der sprachliche Kern der KI . . . . .	105
50.5 Das Problem der Halluzinationen . . . . .	106
50.6 Die Gefahr der Manipulation durch glaubwürdige Fakes . . . . .	106
50.7 Selektive Informationen und die Pluralität der Hintergründe . . . . .	107
50.8 Die Unausweichlichkeit der KI-Entwicklung und die Notwendigkeit der Gestaltung	107
50.9 Weitere Gefahren: Diskriminierung und Überwachung . . . . .	107
50.10 Die Notwendigkeit der Auseinandersetzung mit KI . . . . .	108
<b>51 Beispiele für die Nutzung von Sprachen in der Wissenschaft</b>	<b>109</b>

<b>52 Aufgaben und Fragen, die mit herkömmlichen Methoden nicht lösbar sind</b>	<b>110</b>
52.1 Frage 1: Einfache Aussage in einer Quelle . . . . .	110
52.2 Frage 2: Aussage in Briefen zu einem Thema . . . . .	110
52.3 Frage 3: Aussagen einer Person in ihren Schriften . . . . .	111
52.4 Frage 4: Keine Aussage einer Person in ihren Schriften . . . . .	111
<b>53 Die Herausforderung der inhaltlichen Analyse mit KI</b>	<b>112</b>
53.1 Grenzen der traditionellen Datenbanken . . . . .	112
53.2 Qualifizierte Aussagen auf Basis der verfügbaren Evidenz . . . . .	113
53.3 Herausforderungen bei der Interpretation von Metaphern und Ironie . . . . .	113
53.4 Lernfähigkeit und Entwicklungspotenzial von KI-Systemen . . . . .	114
53.5 Der Paradigmenwechsel durch Large Language Models und Embeddings . . . . .	114
53.6 Die Bedeutung der Philosophie für die KI-Forschung . . . . .	114
<b>54 Traditionell schwer lösbare Fragen in der Forschung</b>	<b>116</b>
54.1 Evidenz finden, um eine Hypothese zu widerlegen . . . . .	116
54.2 Zeitgenössische Autoren und ihre Äußerungen zu historischen Hypothesen . . . . .	116
54.3 Der Einfluss von Publikationen auf historische Autoren . . . . .	117
<b>55 Die Rolle der AI in der geisteswissenschaftlichen Forschung</b>	<b>118</b>
55.1 Die Gefahren der AI und ihre Korrektur durch verbesserte Praktiken . . . . .	118
55.2 Nachvollziehbarkeit von Begründungen für historische Hypothesen . . . . .	118
<b>56 Die Entwicklung der KI und ihr Einfluss auf das wissenschaftliche Arbeiten</b>	<b>120</b>
56.1 Die Entwicklung von Interfaces zur Interaktion mit KI . . . . .	120
56.2 Entwicklung der Eingabemöglichkeiten . . . . .	120
56.3 Der Aufstieg von Chat-GPT . . . . .	121
56.4 Weitere Interaktionsmöglichkeiten . . . . .	121
<b>57 Die Architektur hinter den KI-Systemen</b>	<b>122</b>
57.1 Generative KI . . . . .	122
57.2 Von der Syntax zur Semantik . . . . .	122
57.3 Von der Suche nach Zeichenketten zur Suche nach Inhalten . . . . .	123
57.4 Sätze, Aussagen und Wahrheitswerte . . . . .	123
57.5 Die epistemische Dimension des Wissens . . . . .	124
57.6 Die Maschine lernt, Aussagen zu treffen . . . . .	124
57.7 Die Revolution der Sprachmodelle . . . . .	125
57.7.1 Das Prinzip der Embeddings . . . . .	125
57.7.2 Die Bedeutung eines Satzes . . . . .	125
57.8 Die Herausforderung der Bedeutungsgleichheit . . . . .	126
57.8.1 Beispiele für Bedeutungsgleichheit . . . . .	126
57.9 Die Lösung durch künstliche Intelligenz . . . . .	126
57.9.1 Komplexe Übersetzungen . . . . .	126

57.10 Das Training der KI-Modelle . . . . .	127
57.10.1 Die Parameter der Modelle . . . . .	127
57.10.2 Trainingsdatensätze und Übersetzungsressourcen . . . . .	127
57.11 Weitere Trainingsdaten . . . . .	127
57.12 Kontextabhängigkeit der Bedeutung . . . . .	128
57.13 Funktionsweise der KI bei inhaltlichen Fragen . . . . .	128
57.14 Erweiterung der Embeddings auf multimediale Inhalte . . . . .	128
<b>58 Attention is all you need - die zweite Revolution</b>	<b>129</b>
58.1 Transformation von Sequenzen . . . . .	129
58.2 Die Bedeutung der Sequenztransformation . . . . .	129
<b>59 Rettungsversuch und KI-Demonstration</b>	<b>130</b>
59.1 Die Herausforderung der Feststellung . . . . .	130
59.2 Von der Frage zur Anweisung . . . . .	131
59.3 Die Bedeutung der Aufmerksamkeit . . . . .	131
59.4 Die Macht der Kontextualisierung . . . . .	131
59.5 Das Problem der Halluzination . . . . .	132
59.6 Konsequenzen für die Verwendung von KI-generierten Texten . . . . .	132
<b>60 Erweiterung der KI-Modelle um Wissen und Validierung</b>	<b>133</b>
60.1 Notwendige Ergänzungen für sachliche Korrektheit . . . . .	133
60.2 Verhältnis von Sprache und Sachlichkeit . . . . .	133
<b>61 Auswirkungen der KI-Entwicklung auf Sprache und Bedeutung</b>	<b>134</b>
61.1 Zirkularität der Bedeutung in KI-trainierten Texten . . . . .	134
61.2 Übersetzungsfähigkeiten aktueller Programme . . . . .	134
<b>62 Die Bedeutung der Sprachverwendung</b>	<b>135</b>
<b>63 Die Gefahren fehlerhafter Kontexte</b>	<b>136</b>
<b>64 Erfolge und Anwendungen</b>	<b>137</b>
<b>65 Die Bedeutung des Chats</b>	<b>138</b>
<b>66 Mögliche Projektarbeiten</b>	<b>140</b>
<b>67 Generative Modelle der AI</b>	<b>141</b>
67.1 Vielzahl verschiedener Modelle . . . . .	141
67.2 Funktionsweise der aktuellen Modelle . . . . .	142
67.3 Kommunikation mit KI-Modellen wie mit einer menschlichen Person . . . . .	143
67.4 Stilistische Anpassungsmöglichkeiten . . . . .	143
67.5 Konfiguration formaler inhaltlicher Regeln . . . . .	143

67.6 Charakterdefinition als zusätzliche Dimension . . . . .	144
67.7 Herausforderungen und Zukunftsaufgaben . . . . .	144
67.8 Historisches Schließen als Anwendungsbeispiel . . . . .	144
67.9 Bedeutung des Kontexts . . . . .	145
<b>68 Kontext und Sachkompetenz als Schlüsselfaktoren</b>	<b>146</b>
<b>69 Die Grenzen der KI</b>	<b>147</b>
<b>70 AGI - Der Traum von der Superkompetenz</b>	<b>148</b>
<b>71 Die Zukunft der KI</b>	<b>149</b>
<b>72 Die Bedeutung der Geisteswissenschaften</b>	<b>150</b>
<b>73 Schlussgedanken</b>	<b>151</b>
<b>74 Bewertung der Leistungsfähigkeit von KI-Modellen</b>	<b>152</b>
<b>75 Kontext und Kontextgröße</b>	<b>153</b>
<b>76 Der Nadeltest</b>	<b>154</b>
76.1 Die Eigenschaften einer effektiven Instruktion . . . . .	155
76.2 Von der Query zur Instruktion . . . . .	155
76.3 Die technischen Grundlagen . . . . .	155
<b>77 Die Bedeutung des Kontexts</b>	<b>157</b>
77.1 Eine typische Google-Frage . . . . .	157
77.2 Trainingsgrundlage und Sachkompetenz . . . . .	157
77.3 Grenzen der Sachkompetenz . . . . .	158
<b>78 Die Herausforderung der Aktualität</b>	<b>159</b>
78.1 Die Notwendigkeit kritischer Prüfung . . . . .	159
78.2 Die Problematik widersprüchlicher Informationen . . . . .	160
78.3 Die interne Präferenzordnung der KI-Modelle . . . . .	160
78.4 Die Grenzen der derzeitigen Regeln . . . . .	160
78.5 Die Notwendigkeit hochwertiger Quellen . . . . .	160
78.6 Die Bedeutung von Aktualität und Alter der Quellen . . . . .	161
78.7 Die Sprachkompetenz vs. die Sachkompetenz . . . . .	161
78.8 Die Notwendigkeit eines Austauschs mit anderen Meinungen . . . . .	161
78.9 Die Grenzen der Internetressourcen . . . . .	161
78.10 Die Bedeutung von Meinungsvielfalt und Wahrheit . . . . .	161
78.11 Die Rolle der Wissenschaft . . . . .	162
78.12 Die Herausforderung alternativer Lösungsvorschläge . . . . .	162
78.13 Beispiele für die Kompetenz und Limitierung der Modelle . . . . .	162

78.14 Reformulierung von Fragen zur Präzisierung der Absicht . . . . .	163
78.15 Anreicherung von Instruktionen mit Kontextinformationen . . . . .	163
78.16 Einbeziehung des Dialogkontexts in Chat-Systemen . . . . .	163
78.17 Auflösung von Referenzen durch Kontextberücksichtigung . . . . .	164
78.18 Ausblick: Hybride Modelle der Zukunft . . . . .	164
<b>79 Einstellung eines Konversationsstils</b>	<b>165</b>
<b>80 Grenzen der aktuellen KI-Modelle</b>	<b>166</b>
80.1 Fehlende epistemische Ebene der Prüfung . . . . .	166
80.2 Projekt Lettre AI . . . . .	166
80.2.1 Beispiel 1: Schläger und Ball . . . . .	166
80.2.2 Beispiel 2: Drei Personen im Raum . . . . .	167
<b>81 Philosophie der AI</b>	<b>168</b>
<b>82 Begrüßung und Einführung</b>	<b>169</b>
<b>83 Aktuelle Entwicklungen in der KI</b>	<b>170</b>
<b>84 Erwartungen an KI-Modelle</b>	<b>171</b>
84.1 Generative KI und KI-Charaktere . . . . .	171
84.2 KI - Ein Marketingbegriff? . . . . .	171
<b>85 Anforderungen an zukünftige KI</b>	<b>173</b>
85.1 Defizite aktueller KI-Modelle . . . . .	173
85.2 Unterschied zwischen Information und Wissen . . . . .	174
<b>86 Herausforderungen für die Entwicklung zukünftiger KI</b>	<b>175</b>
86.1 Sprachkompetenz als Kernkompetenz aktueller KI-Modelle . . . . .	175
<b>87 Kontexte und Handlungsanweisungen in der Interaktion mit KI</b>	<b>177</b>
87.1 Instruktionen als Handlungsanweisungen . . . . .	177
87.2 Instruktionen als Kern der KI-Modelle . . . . .	178
<b>88 Lernen von Kompetenz im Hintergrund</b>	<b>179</b>
88.1 Interaktives Lernen durch Nutzerfeedback . . . . .	179
88.1.1 Beispiel: Biografische Informationen zu Leonhard Euler . . . . .	179
88.2 Lernen durch Nutzerdaten und externe Quellen . . . . .	180
88.2.1 Googles Digitalisierungsprojekt mit Bibliotheken . . . . .	180
<b>89 Generierung und Kontext in der Interaktion mit Chatmodellen</b>	<b>181</b>
89.1 Kontextbezogene Antworten . . . . .	181
89.2 Metaebene der Instruktionen . . . . .	181
89.3 Vielschichtigkeit der Kompetenzbereiche . . . . .	181

<b>90 Grenzen aktueller KI-Modelle</b>	<b>183</b>
<b>91 Perspektivwechsel: Erwartungen an eine philosophische KI</b>	<b>184</b>
91.1 Allgemeine künstliche Intelligenz (AGI) . . . . .	184
91.2 Fokus auf spezifische Kompetenzbereiche . . . . .	184
91.3 Semantische Suche . . . . .	185
91.4 Reasoning . . . . .	185
91.5 Die Notwendigkeit der historischen Kontextualisierung . . . . .	185
91.6 Die Herausforderung der Beurteilung historischer Hypothesen . . . . .	186
91.7 Die Bedeutung epistemischer Kompetenz . . . . .	186
<b>92 Die Idee der Individualität von KI-Modellen</b>	<b>187</b>
92.1 Die Voraussetzungen für eine KI-Körperschaft . . . . .	187
<b>93 Historische Vorbilder für die Gestaltung von KI-Modellen</b>	<b>188</b>
93.1 Der vitruvianische Mensch - Die Proportion als Naturgesetz . . . . .	188
93.2 David - Die Freiheit des selbstbestimmten Lebens . . . . .	188
93.3 Der Künstler als Erklärer - Die Fähigkeit zur Rechtfertigung . . . . .	188
<b>94 Das Projekt "Magister AI Faustus"</b>	<b>189</b>
94.1 Kooperation mit der Klassikstiftung Weimar . . . . .	189
94.2 Herausforderung an die gegenwärtigen KI-Modelle . . . . .	189
94.3 Ziel des Projekts . . . . .	190
94.4 Vorgehensweise . . . . .	190
<b>95 Die biografischen Quellen zu Goethes Leben</b>	<b>191</b>
95.1 Die Herausforderung der Kontextualisierung . . . . .	191
95.2 Projektaufgabe und Organisation . . . . .	191
<b>96 Bereiche zukünftiger Kompetenzen von KI-Modellen</b>	<b>193</b>
96.1 Textgenerierung . . . . .	193
96.2 Datenauswertung . . . . .	193
96.3 Ergebniskritik . . . . .	193
<b>References</b>	<b>194</b>

# **Vorlesung *Philosophie der AI***

Diese Website enthält die Living Pages zur Vorlesung *Philosophie der KI* von Prof. Dr. Gerd Graßhoff. Die Vorlesung findet im Sommersemester 2024 an der Humboldt-Universität zu Berlin statt. Die Living Pages der Vorlesung wurden aus dem Transkript der mündlich gehaltenen Vorlesung mit AI transkribiert und mit den Modellen von den Modellen von *Lettre AI* transkribiert und bearbeitet. Die Ergebnisse dieser Transkription werden von Gerd Graßhoff weiter redigiert, ergänzt und mit weiterführenden Links und Verweisen versehen.

Die Seiten richten sich an alle, die sich für die Philosophie der Künstlichen Intelligenz interessieren. Sie stehen unter der Creative Commons Lizenz 4.0 und dürfen gerne zitiert werden. Andere Verwendungen wie auszugsweise Kopien oder digitale Nutzungen erfordern die Erlaubnis des Autors.

# **1 Was ist AI?**

ai\_Vorl1

## 2 Begrüßung und Einführung

Herzlich willkommen zur ersten Vorlesung “Philosophie der AI”! Ursprünglich trug diese Veranstaltung den Titel “Philosophie der künstlichen Intelligenz”, doch angesichts der aktuellen Diskussionen habe ich mich entschieden, den Begriff auf “AI” zu verkürzen. In diesem Semester möchte ich Ihnen einen umfassenden Überblick über die philosophischen Beiträge und Fundamente der modernen Artificial Intelligence geben und Sie durch die Grundlagen führen.

Entgegen der Erwartungen vieler geht es in dieser Vorlesung nicht primär darum, eine Bewertung oder Reflexion über die Folgen und Konsequenzen der künstlichen Intelligenz vorzunehmen. Obwohl wir diese Themen en passant ebenfalls behandeln werden, liegt der Kern der Vorlesung in der Erörterung der Grundthese, dass die eigentliche Innovation und der technologische Kern hinter dem Funktionieren der AI nicht nur in der Informatik, Technologie oder der fortschreitenden Entwicklung der Chips liegt, sondern in der Philosophie selbst. Ich vertrete die Ansicht, dass die künstliche Intelligenz heute eine Renaissance der analytischen Philosophie zur Folge hat, die die eigentliche inhaltliche und systematische Basis dessen bildet, was wir heute unter AI verstehen. Es handelt sich hierbei um eine anspruchsvolle Position, die die Philosophie nicht nur als Kommentator der technologischen und gesellschaftlichen Entwicklungen betrachtet, sondern als essenziellen Teil dieser Bewegung und Entwicklung.

Wir befinden uns derzeit nicht nur inmitten einer technologisch-gesellschaftlichen, politischen und sonstigen Revolution, die in ihrer Tragweite mit der Einführung der Elektrizität vor 150 Jahren oder des Webs vor etwa 25 Jahren vergleichbar ist. Vielmehr stehen wir gerade am Anfang einer Phase der technologischen Revolution durch die Einführung der künstlichen Intelligenz, deren weitreichende Entwicklungen wir nur erahnen können. Ein Indiz dafür ist die Tatsache, dass technologische Veränderungen, Möglichkeiten und Nutzungsformen mittlerweile auf täglicher Basis geschehen.

Während der Vorbereitung dieser Vorlesung ist mir aufgefallen, dass man nicht davon ausgehen kann, mit denselben Utensilien, Tools und Hilfsmitteln zu beginnen und am Ende der Vorlesung weiterzuarbeiten. Die Möglichkeiten und technologischen Anforderungen ändern sich so rasant, dass sie sich sogar während des Verlaufs dieser Vorlesung weiterentwickeln werden. Mein Ziel ist es, Ihnen die Gelegenheit zu bieten, einige dieser Tools während der Vorlesung, in der Nachbereitung oder Vorbereitung selbst auszuprobieren.

## 2.1 Was ist AI?

Künstliche Intelligenz, oder kurz AI, ist ein Begriff für eine technische Möglichkeit, die Mitte der 50er Jahre die Phantasie einer Reihe von Forschern anregte.<sup>1</sup> Diese Phantasien entwuchsen den Arbeiten zu den Grundlagen der Mathematik und Logik, die eine enge Verwandschaft von zahlentheoretischen Fragestellungen mit denen von Algorithmen und der Berechenbarkeit von Problemen betrafen. Alan Turings Arbeiten als Fortsetzung von Kurt Gödels fundamentaler Arbeit über “unentscheidbare Sätze der Principia Mathematica und verwandter Systeme” war der Katalysator für die nachfolgenden Anstrengungen, die theoretischen Möglichkeiten in praktische Anwendungen zu überführen.<sup>2</sup> Ihr Ziel war es, maschinelle Computertechnologien zu entwickeln, die den menschlichen kognitiven Fähigkeiten nicht nur ebenbürtig sind, sondern sie sogar übertreffen. Man versprach damals vollmundig, dass dieses ehrgeizige Ziel in nur drei bis vier Jahren erreicht sein würde. Die Menschheit könnte dann endlich ihre Freizeit in vollen Zügen genießen, nur noch wenige Stunden pro Woche arbeiten, während der Rest von der AI erledigt würde.

Doch wie wir alle wissen, hat sich von dieser Vision bisher nichts eingelöst. Die Vorstellung war, dass AI als Meisterdisziplin des menschlichen Denkens schnell alle Bereiche überflügeln würde. Als Paradebeispiel galt damals das Schachspiel.<sup>3</sup> Doch erst Anfang der 2000er Jahre gelang es einem Computerprogramm, den Schachweltmeister Garri Kasparov in einem ernsthaften Spiel zu besiegen - immerhin 50 Jahre später als ursprünglich prophezeit.

Das andere große Ziel, Computer zu entwickeln, die selbstständig wissenschaftlich kreativ denken können, ist bis heute nicht wirklich erreicht. Trotz aller anderslautenden, manchmal sensationsheischenden Meldungen bin ich jedoch sicher, dass diese Stufe in den nächsten Jahren erreicht werden wird. Dass also wissenschaftliche, kreative, kognitive und intellektuelle Aktivitäten von Maschinen alleine, ohne Assistenz von Forschern gemeistert werden. Das ist sozusagen noch die Krönung der Herausforderung von AI, von Artificial Intelligence.

---

<sup>1</sup>Copeland (2004), [Dartmouth Summer Research Project](#), abgerufen am 15.5.2024.

<sup>2</sup>Turing skizzierte die Grundzüge eines universellen Computers in seiner Vorlesung in der London Mathematical Society 20. Feb 1947. Copeland (2004), S. 378-394. Neumann (1963), Gödel (1986). Von Neumann war tief beeinflusst von Turings Arbeit und setzte sie in der Entwicklung des EDVAC um. Copeland (2004), S. 515.

<sup>3</sup>Samuel (1959), Shannon (1950a), Shannon (1950b), Newborn (1997), Davies (1950)

## 3 AI als Alleskönner

Was Ihnen derzeit tagtäglich in der Öffentlichkeit als AI präsentiert wird, hat mit den eigentlichen Visionen und Zielen oft wenig zu tun. Nehmen wir als Beispiel eine Anzeige der Firma Samsung für ihre "Bespoke AI 11-Kilogramm-Washing-Maschine Serie 8 mit AI-Eco-Bubble und Quick-Drive". Technisch gesehen handelt es sich schlicht um eine Waschmaschine, aber das Label "AI Wash" soll den Verkauf ankurbeln.



Figure 3.1: Samsung AI Waschmaschine

Was ist daran nun wirklich AI? Nicht viel, es ist mehr ein Verkaufsargument als alles andere. Alles, was halbwegs gesteuert ist, wird heutzutage als AI vermarktet. Wenn ich hier "Licht aus" sage und es dunkel würde, würden Sie vielleicht denken "Oh, wir haben AI an der HU". Dabei ist es letztlich nur eine etwas anspruchsvollere Steuerungstechnik, mehr nicht. Das Wort AI ist hier fehl am Platz, auch wenn es gerade überall en vogue ist.

### 3.1 Der Durchbruch der AI-Visionen

Sind wir also jetzt in einer Zeit angekommen, in der sich die ursprünglichen AI-Visionen doch noch erfüllen könnten? Meine Antwort lautet: Ja. Und ich möchte Ihnen heute einen systematischen Grund dafür nennen, der für mich entscheidend ist und den ich Ihnen so vermitteln

möchte, dass er nachvollziehbar wird. Nebenbei bemerkt: Wenn Sie Fragen oder Zwischenfragen haben, melden Sie sich einfach. Dann gestalten wir die Vorlesung etwas lebendiger und interaktiver.

Der Aspekt, auf den ich hinaus möchte und den ich für den Meilenstein halte, ist, dass die AI-Visionen gerade dabei sind Wirklichkeit zu werden. Die AI-Propaganda hingegen, die sollten wir schnell beiseite legen. Das ist in erster Linie ein Verkaufsargument, das nicht den Kern der technologischen Innovation ausmacht. Und genau das soll heute unser Thema sein.

## **3.2 Die Attraktivität von AI**

Wo liegt denn potenziell die Attraktivität der AI, wie immer wir uns ihr auch nähern? Ist es eine bessere Internetsuchmaschine, die derzeit vielleicht eine der Triebfedern ist? Um das zu verstehen, müssen wir uns die Entwicklung des Internets vor Augen führen.

Gemessen an der Technologiegeschichte ist das Internet noch gar nicht so alt, etwas mehr als 20 Jahre. Wer die Anfänge noch miterlebt hat, erinnert sich an die ersten Browser, die damals oft mit Duschanlagen verwechselt wurden. Vor 20 Jahren wussten die wenigsten, was ein Internetbrowser eigentlich ist. Mittlerweile können wir uns ein Leben ohne Internet kaum noch vorstellen, weder technisch noch gesellschaftlich.

## **3.3 Die ursprüngliche Idee des Internets**

Im Kern war die Konstruktion des Internets, die am CERN entwickelt wurde, folgende: Irgendwo stellen wissenschaftliche Einrichtungen webzugängliche Seiten als Informationsquellen bereit. Als Wissenschaftler oder technologische Provider verantworten sie die Inhalte, pflegen sie und sorgen für dauerhafte Zugänglichkeit. Die Browser sind lediglich das lesende Frontend für diejenigen, die auf die Inhalte zugreifen wollen.

Damals war das Internet also eine Art anspruchsvolles Faxgerät als Empfänger der Inhalte. Der Clou lag darin, dass man ganz einfach andere Inhalte per Verlinkung einbinden konnte. So entwickelte sich ein Schneeballsystem, das ein globales Netz von miteinander verknüpften Inhalten erzeugte. Das war die Webrevolution vor 20 Jahren.

## **3.4 Die Ablösung der Webwelt durch AI**

Was wir jetzt erleben, ist eine Ablösung dieser Webwelt durch AI. In den nächsten Monaten werden Sie zunehmend feststellen, dass nicht mehr die Provider die Netzinhalte erstellen, auf Webservern bereitstellen und per Browser zugänglich machen. Diese Grundarchitektur wird abgelöst. Nicht mehr der Browser verantwortet, pflegt und stellt die Inhalte bereit. Das

ist eine revolutionäre Änderung der Architektur der Informationsflüsse, aber auch der damit verbundenen Probleme. Einen Teil davon werden wir noch kennenlernen oder haben Sie schon erfahren.

Das Web funktionierte bisher deshalb, weil die Inhalte von den jeweiligen Personen, Institutionen oder Wissenschaftlern, die sie bereitstellten, auch autorisiert wurden. Für die Korrektheit und Richtigkeit bürgten die Glaubwürdigkeit und Gewissenhaftigkeit der Provider. Das ändert sich jetzt. Und wir alle wissen um die Gefahren, aber auch Potenziale, die damit einhergehen.

- Auf der einen Seite sind es nun große Internetfirmen, die die Inhalte über AI-Maschinen, sogenannte Bots, bereitstellen.
- Auf der anderen Seite können es auch böswillige Gestalten, Institutionen oder Staaten sein, die Inhalte generieren, ins Netz einspeisen, ohne als autorisierende Internetprovider in Erscheinung zu treten.

Derzeit wird das unter dem Stichwort “Internetinhalte der Social Media” diskutiert. Doch das ist nur die Oberfläche. Der Kern des Wandels und des Problems liegt darin, dass die Grundarchitektur des Internets mit den verantwortlichen Providern abgelöst wird durch - ich will nicht sagen unverantwortliche Bots - aber zumindest durch nicht mehr verantwortliche Internetinhaltsprovider. Und das hängt eben mit der AI-Revolution und dem Wandel der Informationsflüsse im Internet zusammen.<sup>#</sup> Die Veränderung der Informationssuche im Zeitalter der Künstlichen Intelligenz

Meine sehr verehrten Damen und Herren, lassen Sie uns heute gemeinsam einen Blick in die Zukunft der Informationssuche werfen. Bislang war es für uns alle selbstverständlich, dass wir bei der Suche nach Informationen auf die Dienste von Suchmaschinen wie Google zurückgreifen konnten. Wir vertrauten darauf, dass die von diesen autoritativen Anbietern bereitgestellten Inhalte glaubwürdig und sorgsam kuratiert waren. Doch in der nächsten Phase der digitalen Revolution wird sich dies grundlegend ändern.

### **3.5 Die Umgestaltung der Architektur des Internets**

Die Architektur des Internets befindet sich in einem extrem dynamischen Wandlungsprozess, dessen Ausgang noch niemand vorhersehen kann. Eines ist jedoch sicher: Es werden enorme Anstrengungen unternommen und gewaltige finanzielle Mittel investiert, um diese Transformation voranzutreiben. Jeder Staat, jede Region und auch Europa sollte ein vitales Interesse daran haben, die Kontrolle über diese Entwicklung nicht zu verlieren.

## **4 Neue Möglichkeiten durch Künstliche Intelligenz**

Doch lassen Sie uns zunächst einen Blick auf die vielversprechenden Möglichkeiten werfen, die uns die Künstliche Intelligenz eröffnet. Vielleicht erscheinen Ihnen einige dieser Anwendungen auf den ersten Blick trivial, doch ich versichere Ihnen, sie haben das Potenzial, unseren Alltag und unsere Arbeit grundlegend zu verändern.

### **4.1 Hochwertige Übersetzungen**

Nehmen wir zum Beispiel das Thema Übersetzungen. Seit Jahrzehnten wurden enorme Ressourcen in die Entwicklung von linguistischen Modellen zur automatischen Übersetzung von Sprachen investiert. Doch lange Zeit waren die Ergebnisse bestenfalls als Partygags zu gebrauchen und keinesfalls für den ernsthaften Einsatz geeignet. In den letzten Jahren hat sich dies jedoch grundlegend geändert. Mittlerweile sind die automatischen Übersetzungen von so hoher Qualität, dass sie sogar für akademische Zwecke genutzt werden können.

Lassen Sie mich Ihnen ein Beispiel aus meinem eigenen Fachgebiet, der Wissenschaftsgeschichte, geben. Viele der historischen Quellen, mit denen wir arbeiten, sind in Latein verfasst. Vor 100 Jahren mussten Doktoranden ihre Dissertationen an unserer Fakultät noch auf Latein einreichen. Heute würden die meisten von Ihnen wohl Schwierigkeiten haben, einen lateinischen Quelltext sinnvoll zu interpretieren. Doch dank der Fortschritte in der Künstlichen Intelligenz gibt es Hoffnung. Vielleicht führen wir ja in unserer Fakultät bald wieder die Pflicht ein, Doktorarbeiten auf Latein zu verfassen - mit AI als Hilfsmittel könnte dies durchaus ein Alleinstellungsmerkmal unserer Universität werden.

### **4.2 Simultanübersetzung und Lektoratsassistentz**

Die Möglichkeiten gehen jedoch noch weiter. In naher Zukunft werden wir in der Lage sein, hervorragende Simultanübersetzungen anzubieten. Ausländische Studierende, die keine europäische Sprache beherrschen, könnten meine Vorlesung mit einem Ohrhörer verfolgen und eine simultane Übersetzung erhalten.

Auch im Bereich des Lektorats gibt es spannende Entwicklungen. Programme wie Grammarly oder DeepL Write bieten bereits heute Textverbesserungsvorschläge, die durchaus mit der Qualität professioneller Lektoratsassistenzen mithalten können. Selbst große wissenschaftliche Verlage wie Nature stellen ihren Autoren mittlerweile Tools zur Verfügung, um ihre englischen Texte in lesbare Form zu bringen. Ob und wie dies gewünscht ist, wird derzeit heiß diskutiert. Doch ich bin davon überzeugt, dass in Zukunft das AI-gestützte Lektorat für wissenschaftliche Publikationen zum Standard werden wird.

### 4.3 Automatisierte Forschungsberichte

Vor der Tür stehen bereits Modelle, die in der Lage sind, eigenständig Texte wie Forschungsberichte zu verfassen. In experimentellen Wissenschaften wie der klinischen Forschung wird bereits daran gearbeitet, Ergebnisse und Erkenntnisse automatisch in Berichte zu überführen, die qualitativ den gängigen Publikationen entsprechen. Dies wirft natürlich Fragen auf:

- Wer ist der Autor eines solchen Berichts?
- Akzeptieren wissenschaftliche Journals Texte, die von einer AI erstellt wurden?
- Wie gehen wir mit Verantwortlichkeit, Seriosität und Zurechenbarkeit um?

Diese Probleme müssen gelöst werden, wenn wir diese Entwicklung weiter vorantreiben wollen.

### 4.4 Das Labor für Lettre AI

In eigenen Labor - *Lettre AI* erforschen und entwickeln wir die hier vorgestellten Techniken um. Unser Ziel ist es, eine AI bereitzustellen, die über die Fähigkeiten des Lesens, Übersetzens und Formulierens hinaus auch epistemische Qualifikationen mitbringt - also wissenbezogene Fähigkeiten, die wir noch näher kennenlernen werden.

Lassen Sie mich Ihnen ein Beispiel für die bereits existierende Leistungsfähigkeit von AI geben. Ich zeige Ihnen hier einen Ausschnitt aus einem Werk, das zu Beginn des 17. Jahrhunderts wie ein Wirbelwind durch Europa fegte: den “Sidereus Nuncius” von niemand geringerem als Galileo Galilei. Dieses Buch markierte den Beginn einer Revolution, denn es war eines der ersten wissenschaftlichen Werke, das nicht nur auf Latein, sondern auch in der Volkssprache Italienisch verfasst wurde und so einer breiteren Öffentlichkeit zugänglich war.

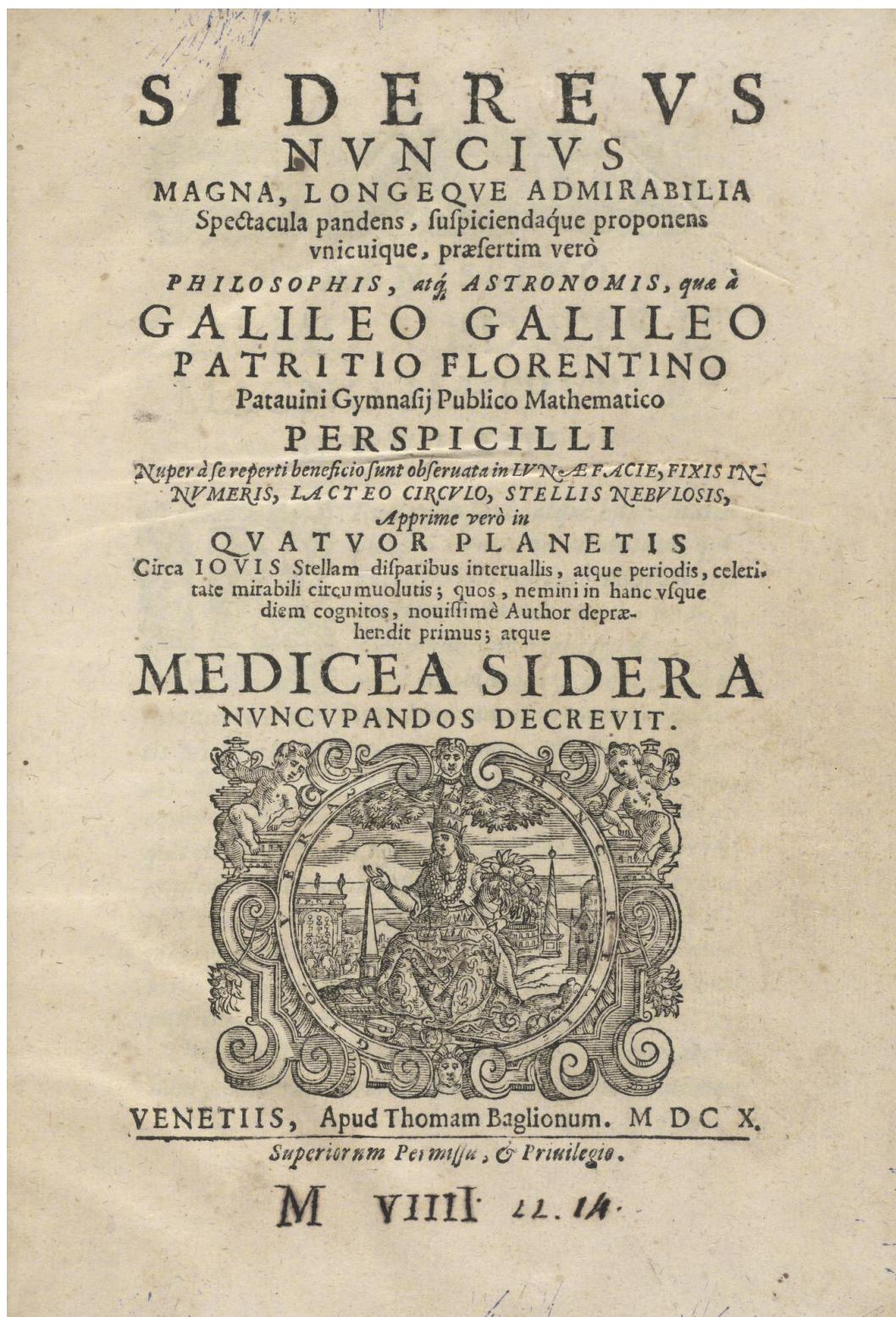


Figure 4.1: Sidereus Nuncius

Ich habe jetzt eine Variante von Chat-GPT aufgebaut. Für diejenigen unter Ihnen, die bereits mit Chat-GPT gearbeitet haben, wird die Oberfläche vertraut aussehen.

## **4.5 Der Kern der Vorlesung**

Doch lassen Sie uns zum Kern dieser Vorlesung kommen. Mir ist es wichtig, dass wir gemeinsam verstehen, was AI eigentlich ist und wie sie funktioniert.<sup>#</sup> Begrüßung und Einführung in Interaktion mit ChatGPT

Stellen Sie sich vor, Sie öffnen die Seite von ChatGPT und werden mit der freundlichen, typisch amerikanischen Begrüßungsfloskel “How can I help you?” empfangen. Es klingt wesentlich servicefreundlicher als ein schlichtes “Hi, hier bin ich”. Die AI bietet Ihnen direkt ihre Hilfe an und präsentiert vier mögliche Optionen, die zwar oft irrelevant sind, Ihnen aber die Mühe ersparen sollen, sich selbst etwas auszudenken. Darunter können Sie dann eingeben, wobei Sie Unterstützung benötigen.

## **4.6 Übertragen eines Bildes in maschinenlesbaren Text**

Nehmen wir an, Sie haben eine Seite mit komplexen Inhalten vor sich, mit denen Sie in ihrer jetzigen Form nichts anfangen können. Hier kommt die AI ins Spiel: Sie können einfach einen Screenshot der Seite machen und diesen in den Chat-GPT hochladen. Anschließend instruieren Sie die AI mit einer Anweisung wie “Transkribiere das Bild” - und schon erhalten Sie eine nahezu fehlerfreie Übertragung des nicht gerade einfachen Textes in getippte Buchstaben. Eine Leistung, die bis heute kein anderes Programm in dieser Qualität vollbringen kann.

## **4.7 Übersetzen des Textes in eine andere Sprache**

Doch das ist erst der Anfang. Nehmen wir an, Sie verstehen kein Latein - kein Problem. Tippen Sie einfach “Übersetze diesen Text ins Deutsche” ein und schon erhalten Sie eine verständliche, wenn auch noch etwas gewöhnungsbedürftige Übersetzung. Mit ein wenig Feinschliff oder dem Wechsel des Modells lässt sich daraus ein publikationsreifer deutscher Text erstellen. Und das Ganze funktioniert nicht nur für Deutsch und Englisch, sondern für über 150 Sprachen weltweit, darunter auch Japanisch und Koreanisch. Selbst obskure mittelalterliche Quellen stellen kein Hindernis dar.

# **5 Erweiterung der Möglichkeiten durch Phantasie und gezielte Fragestellungen**

Doch jetzt fängt der eigentliche Spaß erst an. Mit dem nun zugänglichen Text eröffnen sich ganz neue Möglichkeiten jenseits der typischen Google-Fragen wie "Wer war Galilei?" oder "Wann lebte er?". Stattdessen können Sie die AI mit Fragen herausfordern, die Google unmöglich beantworten kann. Zum Beispiel: "In welcher Stadt trank Galilei im Mai 1615 ein Glas Wein?". Das Problem liegt hier nicht nur darin, dass Google dieses spezifische Ereignis nicht kennt, sondern dass eine einfache Stichwortsuche prinzipiell nicht ausreicht, um die Antwort zu finden.

## **5.1 Analogie zu Sherlock Holmes**

Stellen Sie sich die AI als eine Art elektronischen Sherlock Holmes vor. Sie nimmt das gesamte Universum an Dokumenten über Galilei zur Kenntnis - seine Briefe, seine historischen Lebensumstände, seine typischen Aktivitäten im Frühjahr 1605. Aus diesen Informationen zieht sie dann Rückschlüsse und generiert eine fundierte Hypothese darüber, wo und wann Galilei wahrscheinlich sein Glas Wein genossen hat. Zwar nicht mit absoluter Sicherheit, aber basierend auf seinen regelmäßigen Lebensumständen. Solche Fragen werden die AI-Modelle in naher Zukunft beantworten können.

## **5.2 Vielfältige Analysemöglichkeiten von Texten**

Doch damit nicht genug. Sie können die AI auch anweisen, eine Tabelle mit allen Verben des Textes zu erstellen oder gezielt nach Verben zu suchen, die ein Lob, eine Ankündigung oder ein Versprechen ausdrücken - selbst wenn Sie die genaue Formulierung nicht kennen. Die Möglichkeiten sind schier grenzenlos.

Ein konkretes Beispiel: Fragen wir die AI, wer sich laut dem Text bewegt. Nach kurzer Bedenkzeit liefert sie die korrekte Antwort: Die vier Planeten bewegen sich zu verschiedenen Zeiten und mit erstaunlicher Geschwindigkeit um den Stern Jupiter - eine Entdeckung, die Galilei machte und die tatsächlich im lateinischen Originaltext erwähnt wird.

# **6 Philosophie als Grundlage für die Möglichkeiten der AI**

Doch wie ist das alles möglich? Die Antwort liegt in der Philosophie - nicht in der Technik. Natürlich brauchen wir auch die technische Infrastruktur, so wie wir Beamer und Notebooks benötigen. Aber der eigentliche Schlüssel zu den Fähigkeiten der AI ist philosophischer Natur. Das wird oft übersehen, doch ich möchte Ihnen zeigen, warum Philosophie hier so entscheidend ist.

## **6.1 Beantwortung von Fragen über Mikrofoneingabe**

Um das Potenzial der AI weiter zu verdeutlichen, können wir auch das Mikrofon aktivieren und eine Frage stellen: "Hat Galilei diese Entdeckung selbst durch Beobachtungen gemacht?". Das System denkt kurz nach und liefert dann die zutreffende Antwort: Ja, laut den Angaben im Text hat Galilei die Entdeckung tatsächlich selbst durch Beobachtungen gemacht.

Das Erstaunliche daran ist nicht nur, dass überhaupt eine Antwort generiert wird, sondern vor allem die Qualität dieser Antwort - trotz Versprechern und spontaner Formulierung mein-erseits.# Einführung in die sprachliche Dimension der AI

Meine Damen und Herren, heute möchte ich Ihnen eine faszinierende und zugleich beunruhigende Entwicklung in der Welt der künstlichen Intelligenz näherbringen. Es geht um die Fähigkeit von AI-Systemen, nicht nur Informationen aus autoritativen Quellen zu sammeln, sondern eigenständig Antworten zu generieren und Inhalte zu erstellen. Diese Entwicklung hat weitreichende Konsequenzen für unser Verständnis von Wissen und Informationsverarbeitung.

## **6.2 Die Möglichkeiten der AI**

Die Möglichkeiten der AI sind atemberaubend und erweitern sich täglich. Lassen Sie mich Ihnen einige Beispiele nennen:

- Übersetzung: AI-Systeme können Texte von einer Sprache in eine andere übersetzen, und zwar mit einer Genauigkeit und Geschwindigkeit, die menschliche Übersetzer in den Schatten stellt.

- Bild-zu-Text-Konvertierung: AI kann Bilder analysieren und deren Inhalt in Textform beschreiben. Dies eröffnet völlig neue Möglichkeiten der Bildverarbeitung und -archivierung.
- Audio-zu-Text-Konvertierung: Gesprochene Sprache kann von AI-Systemen in Echtzeit transkribiert werden, was die Erstellung von Protokollen und Untertiteln erleichtert.
- Textzusammenfassung: Geben Sie der AI ein ganzes Buch, und sie wird Ihnen eine prägnante Zusammenfassung liefern. Dies kann die Recherche und das Studium enorm beschleunigen.
- Text-zu-Audio-Konvertierung: Umgekehrt kann AI auch geschriebenen Text in gesprochene Sprache umwandeln, was neue Möglichkeiten für Hörbücher und Sprachassistenten eröffnet.
- Text-zu-Video-Konvertierung: Hier wird es geradezu unheimlich. AI kann aus Textbeschreibungen realistische Videos generieren, die kaum noch von echten Aufnahmen zu unterscheiden sind.

## 6.3 Gefahren der AI

So faszinierend diese Möglichkeiten auch sind, sie bergen auch erhebliche Risiken. Ein zentrales Problem ist das Phänomen der “Halluzination”. Dabei generiert die AI scheinbar plausible Informationen, die jedoch nicht der Realität entsprechen.

Ein Beispiel: Ich fragte eine AI nach dem Namen der zweiten Frau des Mathematikers Leonhard Euler. Die Antwort klang überzeugend, inklusive eines Verweises auf eine Publikation der Petersburger Akademieschriften von 1784. Doch diese Publikation existiert gar nicht, und die genannte Person war nie mit Euler verheiratet.

Solche Halluzinationen können fatale Folgen haben, wenn sie unerkannt bleiben. Wer eine solche Information zitiert, disqualifiziert sich wissenschaftlich für immer. Dieses Problem trat auch bei der Mars-Mission der NASA auf, als eine AI falsche Informationen über einen Erkundungssatelliten verbreitete.

## 6.4 Der sprachliche Kern der AI

Bei all diesen Anwendungen, sei es Bild-, Audio- oder Videoverarbeitung, bildet die Sprache den Kern der AI-Technologie. Selbst bei der Bildanalyse übersetzt die AI zunächst das Bild in eine verbale Beschreibung, bevor sie weiterverarbeitet wird.

Diese Erkenntnis ist philosophisch bedeutsam und erinnert an Wittgensteins These von der Unhintergehrbarkeit der Sprache. Die sprachliche Verbalisierung von Inhalten ist der Dreh- und Angelpunkt der AI, und genau darum soll es in dieser Vorlesung gehen.

Ich werde mich nicht auf die technischen Details der AI-Entwicklung konzentrieren, sondern auf den Umgang mit Sprache in AI-Modellen. Die anderen Medien sind zwar faszinierend, aber letztlich sekundär. Unser roter Faden wird die philosophische Dimension der sprachlichen Verarbeitung in der AI sein. # Gefahren und Probleme der künstlichen Intelligenz

Meine Damen und Herren, lassen Sie uns heute über die Schattenseiten der künstlichen Intelligenz sprechen. Wir haben bereits die atemberaubenden Möglichkeiten dieser Technologie gesehen, doch nun ist es an der Zeit, auch die Probleme und Gefahren zu beleuchten, die damit einhergehen.

## 6.5 Das Problem der Halluzinationen

Eines der ersten Probleme, auf das wir stoßen, sind die sogenannten Halluzinationen der AI-Modelle. Ein eindrucksvolles Beispiel dafür lieferte das Supermodell von Google, das auf die Frage "Wer fliegt denn da?" eine Antwort gab, die zwar plausibel klang, aber rein fiktiv war. Ohne Zugriff auf aktuelle NASA-Informationen oder Tagesnachrichten erfand das Modell kurzerhand einen Satellitennamen. Innerhalb einer halben Stunde wurde es vom Netz genommen, und der Marktwert von Google-Aktien sank um Millionen. Seitdem trauen sich die Unternehmen nicht mehr, ihre Modelle zu veröffentlichen.

Doch warum halluzinieren die Modelle überhaupt, wenn sie doch schon so viele Fähigkeiten besitzen? Die Antwort darauf ist komplexer als man denkt.

## 6.6 Die Gefahr der Manipulation durch glaubwürdige Fakes

Ein weiteres Problem, das eng mit den Halluzinationen verbunden ist, ist die Fähigkeit der AI, glaubwürdige Texte, Bilder und sogar Videos zu produzieren. Dies öffnet Tür und Tor für falsche oder manipulative Informationen, die auf den ersten Blick echt erscheinen.

Ein aktuelles Beispiel dafür sind die Videos, die im Zusammenhang mit dem Raketenüberfall auf Israel in den sozialen Medien aufgetaucht sind. Sie zeigten panische Einwohner von Tel Aviv, die vor nicht existierenden Einschlägen flohen. Diese Videos wurden absichtlich generiert, um die Öffentlichkeit zu täuschen, und sind für den Betrachter zunächst nicht als Manipulation zu erkennen.

## **6.7 Selektive Informationen und die Pluralität der Hintergründe**

Jede Antwort, die uns ein AI-Modell gibt, basiert auf bestimmten Annahmen und Voraussetzungen. Diese haben jedoch immer auch Alternativen, die möglicherweise nicht besser oder schlechter sind, aber eine Pluralität an Hintergründen darstellen.

Wenn wir eine bestimmte Antwort akzeptieren, akzeptieren wir auch die Voraussetzungen dafür und vernachlässigen die Alternativen. Ein Beispiel dafür ist die Anfrage an ein AI-Modell, ein Porträt eines möglichen Nachfolgers des jetzigen Papstes zu erstellen. Aufgrund der politisch korrekten Voreinstellung des Modells wurde eine farbige Frau im Papstgewand generiert - eine Darstellung, die in der Realität aufgrund der Zusammensetzung des Kardinalskollegiums höchst unwahrscheinlich ist.

Dieses Beispiel verdeutlicht, wie selektive Informationen zu verzerrten Ergebnissen führen können. Es wirft die Frage auf, wie wir mit diesen Problemen umgehen sollen.

## **6.8 Die Unausweichlichkeit der AI-Entwicklung und die Notwendigkeit der Gestaltung**

Eines ist klar: Wir können uns vor diesen Fragen nicht drücken. Die Entwicklung der künstlichen Intelligenz ist unwiderstehlich und unausweichlich. Ab heute werden uns diese Technologien mit all ihren Vor- und Nachteilen zunehmend beschäftigen.

Wir müssen lernen, damit umzugehen und die Entwicklung aktiv mitzugestalten. Nicht im Sinne einer Kontrolle, sondern einer Gestaltung. Denn wenn wir jetzt nicht eingreifen, laufen wir Gefahr, die Kontrolle über diesen Prozess zu verlieren.

## **6.9 Weitere Gefahren: Diskriminierung und Überwachung**

Neben der selektiven Information gibt es weitere Gefahren, die wir im Auge behalten müssen. Dazu gehören Dimensionen der Diskriminierung, bei denen bestimmte Personengruppen oder Qualifikationen berücksichtigt werden, andere hingegen nicht.

Auch die Möglichkeiten der Überwachung durch AI-Systeme sind alarmierend. Ein Beispiel dafür ist China, wo Besucher bei der Einreise lediglich in eine Kamera lächeln müssen und dann während ihres gesamten Aufenthalts live verfolgt und protokolliert werden.

Diese Entwicklungen werfen Fragen auf, wie weit solche Technologien zugelassen und kontrolliert werden sollten. Eine Antwort darauf zu finden, ist keine leichte Aufgabe.

## 6.10 Die Notwendigkeit der Auseinandersetzung mit AI

Angesichts dieser erschütternden Probleme könnte man geneigt sein, das Thema AI einfach zu vergessen. Wozu sich mit Übersetzungen von Galileis lateinischen Texten beschäftigen, wenn wir dafür doch unsere Gelehrten haben?

Doch so einfach ist es nicht. Die Vorteile der künstlichen Intelligenz sind zu groß, um sie zu ignorieren. Wir müssen uns mit dieser Technologie auseinandersetzen, ihre Möglichkeiten nutzen und gleichzeitig ihre Schattenseiten im Blick behalten. Nur so können wir eine Zukunft gestalten, in der die AI zum Wohle der Menschheit eingesetzt wird. # Begrüßung und Einführung

Einen schönen guten Tag, meine Damen und Herren. Heute möchte ich mit Ihnen über zwei Fragen sprechen, die mir in letzter Zeit immer wieder begegnen. Zunächst einmal habe ich eine Frage zu der Konferenz, von der ich gehört habe, dass sie in diesem Monat August stattfinden soll. Wo genau findet diese Konferenz statt?

Ah, ich verstehe. Es handelt sich also um eine regelmäßig wiederkehrende Konferenzserie, die im August abgehalten wird. Es ist bemerkenswert, wie weit fortgeschritten die Aufmerksamkeit und das Wissen um diese Themen inzwischen auch in den Institutionen sind. Sogar auf EU-Ebene wurde im März letzten Jahres bereits ein Bericht veröffentlicht, in dem diese Angelegenheiten thematisiert wurden. Allerdings wurden sie dort so behandelt, wie ich es gerade geschildert habe - sie wurden angesprochen, aber nicht gelöst.

Es gibt keine einzelne Konferenz, die sich des Problems annimmt und von der wir erwarten könnten, dass es in kürzester Zeit gelöst wird. Nein, so einfach ist es leider nicht. Stattdessen möchte ich auf die philosophischen Aspekte eingehen, die sowohl den Vorteilen als auch den Gefahren zugrunde liegen.

## 7 Nutzungsmöglichkeiten in der Wissenschaft

Lassen Sie uns ein Beispiel betrachten, das wir gerade schon diskutiert haben. In der Fachliteratur hält sich hartnäckig das Gerücht, dass Galileis Vater sich negativ über die wissenschaftliche Nutzung anderer Sprachen als Latein geäußert haben soll. Das würde natürlich einen spannenden Vater-Sohn-Konflikt darstellen, denn Galilei selbst ist ja berühmt dafür, dass er das Italienische für die Wissenschaft nutzbar machte, indem er auf Italienisch publizierte.

In zahlreichen Sekundärquellen findet man die These, dass sein Vater dies nicht für wissenschaftlich hielt und dass sein Sohn Galileo Galilei sich besser von diesen italienischen Publikationen fernhalten sollte. Oh, Moment mal - da steht, dass Kepler sich gegenüber Galilei negativ geäußert hat, nicht Galileis Vater. Danke für den Hinweis! Das ist keine Halluzination, sondern ein echter Fehler meinerseits. Ich hoffe, ich vergesse nicht, das für die Internetversion zu korrigieren.

Die Pointe ist jedenfalls, dass man eine solche Frage - ob sich eine Person X irgendwo negativ zu einer bestimmten These geäußert hat - mit Google nicht beantworten kann. Das mag trivial klingen, aber im Moment ist es tatsächlich nicht möglich, dies durch eine Google-Suche herauszufinden. Warum? Weil Google Ihnen kein Dokument im Internet liefern wird, in dem diese Frage direkt beantwortet wird. Und wenn es ein solches Dokument nicht gibt, ist die Frage für Sie mit Google-Techniken nicht zu beantworten.

Dabei handelt es sich um eine Frage, die historisch gesehen entweder wahr oder falsch ist. Wie kann man das also entscheiden? Nicht mit den heutigen Google-Techniken. Hier braucht es eine neue Dimension der Recherche, die über bestimmte Fähigkeiten verfügen muss.

# **8 Bislang nicht lösbarer Aufgaben**

Lassen Sie mich Ihnen anhand einer Liste von Aufgaben und Fragen veranschaulichen, wie zunehmend Probleme auftauchen, die mit den heutigen akademischen Techniken nicht zu lösen sind. Ich spreche hier von Fragen, die selbst Sie als forschende Person nicht beantworten können, wenn sie halbwegs komplex sind.

Mir geht es um die unlösbaren Probleme der realen Forschungswelt, die zwar mit AI lösbar wären, aber aufgrund bestimmter fehlender Fertigkeiten bisher nicht gelöst werden können. Jetzt befinden wir uns im philosophischen Teil meiner Ausführungen und ich werde versuchen, dies sprachanalytisch zu komprimieren.

## **8.1 Frage 1: Einfache Aussage in einer Quelle**

Angenommen, Person A äußert sich in einer Quelle Q zu einer Person namens Jochen Schmidt. Ist diese Aussage wahr oder falsch? Hier haben Sie noch eine gewisse Chance, die Frage eindeutig zu beantworten, wenn Sie die Quelle Q gefunden haben und darin die Person A benannt wird und sich zu Jochen Schmidt äußert. Der Anforderungsgrad ist hier noch nicht sehr hoch. Wenn das Ihre Examensaufgabe wäre, hätten Sie eine realistische Chance, sie zu lösen. Sie müssten nur so lange alle Quellen durchlesen, bis Sie die richtige gefunden haben.

## **8.2 Frage 2: Aussage in Briefen zu einem Thema**

Nehmen wir an, Person A äußert sich in ihren Briefen zu einem Thema T. Das können Sie schon nicht mehr ohne weiteres lösen, ohne eine Lebensdauer damit zu verbringen, das gesamte Schrifttum von Person A zu lesen. Wenn Sie z.B. für eine Examensarbeit eine Biografie über eine Person namens Heinz Müller verfassen sollten und eine solche Aufgabe hätten, müssten Sie zunächst alle Briefe zusammentragen und sie komplett lesen. Und selbst dann wären Sie sich nicht sicher, ob Sie wirklich alle Briefe gefunden haben.

Denken Sie nur an die Kafka-Forscher. Wenn Sie wissen wollen, ob sich Kafka in seinen Briefen jemals zu einem bestimmten Thema geäußert hat oder nicht, haben Sie einen enormen manuellen Forschungsaufwand vor sich, um überhaupt in die Nähe einer Antwort zu kommen. Hier befinden wir uns bereits in Bereichen, die schwer zu beantworten sind - Fragestellungen, die bislang praktisch nicht zu lösen waren.

### **8.3 Frage 3: Aussagen einer Person in ihren Schriften**

Hat eine Person A in ihren Schriften Aussagen der Art T getroffen, wenn Person A sehr viel geschrieben hat? Nehmen wir als Beispiel die Briefe Napoleons. Hat sich Napoleon jemals zu Aspekten der Vorläufer der Genfer Konvention bei der Kriegsführung geäußert? Das können Sie aus praktischen Gründen nicht lösen. Ich will an dieser Stelle nicht sagen, dass es prinzipiell unmöglich ist, aber in der Wissenschaft möchte man solche Fragen beantwortet haben. Und das gilt nicht nur für das öffentliche Interesse, sondern auch für die Wissenschaft selbst.

Sie können sich vorstellen, welch enorme Konsequenzen es für die Wissenschaft hätte, wenn man solche Fragen überhaupt beantworten könnte. Dann wäre es möglich, weitreichende Thesen zu Napoleons Verständnis von Krieg und Frieden aufzustellen, die von der Evidenz abhängen, mit der man solche Fragen beantworten kann. Im Moment ist das nicht möglich.

### **8.4 Frage 4: Keine Aussage einer Person in ihren Schriften**

Angenommen, Person A hat in ihren Schriften keine Aussage T getroffen. Als normaler arbeitender Historiker oder Geisteswissenschaftler werden Sie diese Frage nicht seriös beantworten können. Deshalb gibt es in der Literatur die Unsitten, andere Werke zu zitieren, die sich aus irgendwelchen Gründen dazu bemüßt fühlten, solche Fragen zu beantworten.

Ein Beispiel: Nehmen wir wieder Kafka. Manche Autoren vertreten die These, dass Kafka sich nie antisemitisch geäußert hat. Aber welche Evidenz können Sie dafür eigentlich angeben? Es ist schwierig, eine nicht vorhandene Lektüre von Briefen als Beleg anzuführen. Wie wollen Sie eine solche These rechtfertigen, wenn Sie sie vertreten?

Eine der größten Unsitten der gegenwärtigen akademischen Literatur besteht darin, nicht selbst das Risiko einer These einzugehen, sondern stattdessen den berühmten Heinz Müller zu zitieren, weil er schon einmal etwas Ähnliches gesagt hat. Also fügt man eine Fußnote in die Arbeit ein: "Heinz Müller, 1973, Seite 5: Ganz klar, Kafka hat sich nie antisemitisch geäußert." Und auf einmal entsteht ein Schneeballsystem, das dem Halluzinationseffekt ähnelt, den wir gerade hier hatten. Und zwar nur deshalb, weil die Evidenz, die für bestimmte Thesen erforderlich ist, auf manuelle Weise kaum zu beschaffen ist. Mit AI werden Sie das in Zukunft können.

# **9 Die Herausforderung der inhaltlichen Analyse mit AI**

Jetzt werden Sie vielleicht fragen: Inwiefern ist das speziell für AI relevant? Man könnte doch erwarten, dass sich das grammatisch lösen lässt. Wenn ich die Aussage T formalisieren kann, müsste ich doch auf dem Textkorpus einfach prüfen können, ob diese Bedingung irgendwo erfüllt ist, oder?

Genau das ist der springende Punkt, und ich muss jetzt ein bisschen auf die Uhr schauen, damit ich meine Kurve hier noch hinbekomme. Aber diese Kurve berührt schon das Thema. Was heißt es, in Ihrem Korpus prüfen zu können?

Nehmen wir an, Sie hätten den Idealfall: Kafkas gesammelten Briefwechsel in einer Datenbank. Jetzt möchten Sie wissen, ob es darin eine antisemitische Formulierung gibt. Wie sieht die denn aus? Wenn Sie Ihre Datenbank nach Art einer Google-Suche nach bestimmten Wortvorkommnissen durchforsten, dann können Sie das lösen. Das ist die klassische Vorgehensweise.

Aber inhaltlich betrachtet: Was ist eigentlich eine antisemitische Äußerung? Sobald es darum geht - und deshalb habe ich es hier erwähnt - kön# Betrachtungen zur künstlichen Intelligenz und Sprachverarbeitung

Meine sehr geehrten Damen und Herren, liebe Studierende,

in der heutigen Vorlesung möchte ich Ihnen einen faszinierenden Einblick in die Welt der künstlichen Intelligenz und insbesondere deren Fähigkeiten zur Sprachverarbeitung geben. Wir werden uns mit der Frage beschäftigen, inwieweit AI-Systeme in der Lage sind, komplexe sprachliche Konstrukte wie Metaphern, Ironie oder versteckte Bedeutungen zu erkennen und zu interpretieren.

## **9.1 Grenzen der traditionellen Datenbanken**

Zunächst einmal möchte ich klarstellen, dass ich keineswegs behauptet habe, es gäbe in den vorliegenden Dokumenten keine relevanten Satzvorkommnisse. Die herkömmliche Art der Dokumentenaufzeichnung und -abfrage, wie sie etwa mit Datenbanken möglich ist, erlaubt zwar das Auffinden bestimmter Textpassagen, jedoch keine inhaltlichen Suchen im eigentlichen Sinne.

Selbst moderne AI-Systeme können nicht mit absoluter Sicherheit feststellen, dass eine bestimmte Aussage nicht getroffen wurde, da stets die Möglichkeit besteht, dass die zugrunde

liegende Datenbasis unvollständig ist. Vielmehr lässt sich hier nur mit Wahrscheinlichkeiten operieren - ein Begriff, den ich an dieser Stelle allerdings kritisch hinterfragen möchte.

## **9.2 Qualifizierte Aussagen auf Basis der verfügbaren Evidenz**

Wahrscheinlichkeiten sind numerische Werte zwischen 0 und 1, die man in diesem Kontext nicht sinnvoll einsetzen kann. Stattdessen sollte man sich auf die konkrete Situation beziehen und feststellen: Auf Basis dieser und jener Grundgesamtheit von Briefwechseln und Äußerungen, die als Dokumente für die Befunde zur Verfügung stehen, lässt sich unter der Voraussetzung, dass sie die alleinige Entscheidungsgrundlage bilden, folgendes Fazit ableiten.

Eine solche differenzierte Betrachtung der Befundlage ist unerlässlich, denn es lässt sich ja nicht ausschließen, dass genau jene Briefe, die möglicherweise relevante Inhalte enthalten, vernichtet wurden. Ein solches Szenario würde den Wahrheitswert der Fragestellung grundlegend verändern. Auch AI-Systeme können diese Problematik nicht vollständig ausräumen, sehr wohl aber eine qualifizierte, auf der verfügbaren Evidenz basierende Antwort geben.

## **9.3 Herausforderungen bei der Interpretation von Metaphern und Ironie**

Ein besonders spannendes Feld ist die Fähigkeit von AI-Systemen, mit Metaphern und ungewöhnlichem Sprachgebrauch umzugehen. Gerade im Kontext des Antisemitismus verbergen sich oft codierte Botschaften hinter scheinbar harmlosen Formulierungen. Während eine Blut- und Boden-Ideologie relativ leicht zu identifizieren ist, stellt die Interpretation von Begriffen wie "entwurzelt" oder "ohne Verwurzelung" eine ungleich größere Herausforderung dar.

Anhand eines konkreten Beispiels möchte ich Ihnen verdeutlichen, wozu moderne AI-Systeme in diesem Bereich bereits in der Lage sind. In München hatten wir es mit revolutionären Briefen aus der Zeit der Französischen Revolution zu tun, die in elegantem Französisch verfasst waren und vor Ironie und Sarkasmus nur so strotzten. Um diese Feinheiten zu erkennen, bedarf es zunächst einmal exzellenter Sprachkenntnisse. Doch selbst dann gilt es, die ironischen Komponenten als solche zu identifizieren.

Ich kann Ihnen versichern, dass AI-Systeme mittlerweile über eine Sprachkompetenz verfügen, die es ihnen erlaubt, auch diese Dimension der Sprachverwendung zu erkennen. Allerdings dürfen Sie sich das nicht als simples Schwarz-Weiß-Schema vorstellen, bei dem man einfach einen "Ironie-Kompetenz-Knopf" umlegt und schon funktioniert alles wie bei einem literarischen Meisterinterpreten.

## **9.4 Lernfähigkeit und Entwicklungspotenzial von AI-Systemen**

Vielmehr müssen Sie sich den Lernprozess der AI ähnlich vorstellen wie Ihre eigene Entwicklung zu Beginn Ihres Studiums. Auch Sie haben im Laufe der Zeit eine Menge dazugelernt und sich weiterentwickelt. Genauso können auch AI-Modelle lernen und sich verbessern. Ich möchte keineswegs behaupten, dass bereits alle Probleme und Herausforderungen gelöst sind, aber es gibt vielversprechende Lösungsansätze, um auch mit komplexeren Formen der Sprachverwendung umgehen zu können.

In München haben wir beispielsweise erfolgreich getestet, ob AI-Systeme in der Lage sind, bissige Karikaturen aus den 1920er Jahren zu interpretieren und zu erkennen, welche Personen mit welchen Klischees auf den Arm genommen werden. Mit dem richtigen Training ist es den Bilderkennungsalgorithmen tatsächlich gelungen, diese Zusammenhänge zu entschlüsseln.

## **9.5 Der Paradigmenwechsel durch Large Language Models und Embeddings**

Der entscheidende Unterschied und gleichzeitig der Punkt, an dem der “Philosophical Turn” der AI einsetzt, liegt in der Entwicklung von Techniken wie Large Language Models oder Embeddings. Diese ermöglichen eine Abkehr von der reinen Textsuche hin zu einer inhaltlichen Erfassung der Bedeutung sprachlicher Ausdrücke. Dieser semantische Wechsel, den ich auch gerne als “Semantic Turn” bezeichne, ist der Schlüssel zu den beeindruckenden Fähigkeiten moderner AI-Systeme.

Egal ob es um die Analyse von Bildern, Texten oder Audioaufnahmen geht - all diesen Anwendungen liegt zugrunde, dass die Systeme nicht nur nach bestimmten Zeichenfolgen suchen, sondern deren Bedeutung erfassen und identifizieren können. Genau darum geht es bei den milliardenschweren Investitionen in diesem Bereich: den Modellen beizubringen, auf Basis der eingegebenen Daten die dahinterstehende Semantik zu erkennen.

## **9.6 Die Bedeutung der Philosophie für die AI-Forschung**

Damit eröffnet sich ein weites Feld für die Philosophie. Solange wir nur von Sätzen sprechen, bewegen wir uns auf der Ebene von Formulierungen und syntaktischen Strukturen. Wenn wir jedoch nach der Bedeutung eines Ausdrucks fragen, betreten wir Neuland. Genau hier setzt die aktuelle AI-Revolution an, und deshalb ist die Philosophie von zentraler Bedeutung für diese Entwicklung.

Als Studierende der Philosophie sollten Sie mit der klassischen Unterscheidung zwischen Satz und Aussage vertraut sein. Im Deutschen ist diese Differenzierung von größter Wichtigkeit, während sie in englischen Übersetzungen oft vernachlässigt wird. So haben etwa die Übersetzer

von Wittgensteins Gesammelten Werken sowohl für “Aussage” als auch für “Satz” durchgängig den Begriff “Sentence” verwendet, was zu erheblichen Missverständnissen führen kann. Im Englischen heißt es korrekterweise “Sentence” für Satz und “Proposition” für Aussage.

Genau diese Unterscheidung markiert die fundamentale Revolution, die sich gerade vollzieht: Wir haben es nun mit Maschinen zu tun, die mit Aussagen umgehen können. Und nur Aussagen, nicht Sätze, können wahr oder falsch sein. Wer also über Fake News, Halluzinationen und ähnliche Phänomene spricht und sich dabei auf Sätze bezieht, liegt philosophisch gesehen völlig falsch. Wahrheit und Falschheit können sich konzeptionell nur auf Aussagen beziehen.

Die Tatsache, dass AI-Systeme nun in der Lage sind, sich mit Aussagen zu befassen, birgt ebenso faszinierende Möglichkeiten wie Gefahren. In der nächsten Vorlesung werden wir uns eingehender mit diesen Aspekten beschäftigen und uns ansehen, wie genau diese neuen Technologien funktionieren und welche Auswirkungen sie haben können.

# **10 Philosophie der AI**

ai\_Vorl2

# **11 Begrüßung und Einführung in die Vorlesung “Philosophie der AI”**

Herzlich willkommen zur ersten Vorlesung “Philosophie der AI”! Ursprünglich trug diese Veranstaltung den Titel “Philosophie der künstlichen Intelligenz”, doch angesichts der aktuellen Diskussionen habe ich mich entschieden, den Begriff auf “AI” zu verkürzen. In diesem Semester möchte ich Ihnen einen umfassenden Überblick über die philosophischen Beiträge und Fundamente der modernen Artificial Intelligence geben und Sie durch die Grundlagen führen.

## **11.1 Die Rolle der Philosophie in der AI**

Entgegen der Erwartungen vieler geht es in dieser Vorlesung nicht primär darum, eine Bewertung oder Reflexion über die Folgen und Konsequenzen der künstlichen Intelligenz vorzunehmen. Obwohl wir diese Themen en passant ebenfalls behandeln werden, liegt der Kern der Vorlesung in der Grundthese, dass die eigentliche Innovation und der technologische Kern hinter dem Funktionieren der KI nicht nur in der Informatik, Technologie oder der fortschreitenden Entwicklung der Gerätschaften und Chips liegt, sondern in der Philosophie selbst. Ich vertrete die Ansicht, dass die künstliche Intelligenz heute eine Renaissance der analytischen Philosophie zur Folge hat, die die eigentliche inhaltliche und systematische Basis dessen bildet, was wir heute unter KI verstehen. Es handelt sich hierbei um eine anspruchsvolle Position, die die Philosophie nicht nur als Kommentator der technologischen und gesellschaftlichen Entwicklungen betrachtet, sondern als essenziellen Teil dieser Bewegung und Entwicklung.

# **12 Die KI-Revolution und ihre Auswirkungen**

## **12.1 Eine technologisch-gesellschaftliche Revolution**

Wir befinden uns derzeit nicht nur inmitten einer technologisch-gesellschaftlichen, politischen und sonstigen Revolution, die in ihrer Tragweite mit der Einführung der Elektrizität vor 150 Jahren oder des Webs vor etwa 25 Jahren vergleichbar ist. Vielmehr stehen wir gerade am Anfang einer Phase der technologischen Revolution durch die Einführung der künstlichen Intelligenz, deren weitreichende Entwicklungen wir nur erahnen können. Ein Indiz dafür ist die Tatsache, dass technologische Veränderungen, Möglichkeiten und Nutzungsformen mittlerweile auf täglicher Basis geschehen.

## **12.2 Herausforderungen in der Vorlesungsvorbereitung**

Während der Vorbereitung dieser Vorlesung ist mir aufgefallen, dass man nicht davon ausgehen kann, mit denselben Utensilien, Tools und Hilfsmitteln zu beginnen und am Ende der Vorlesung weiterzuarbeiten. Die Möglichkeiten und technologischen Anforderungen ändern sich so rasant, dass sie sich sogar während des Verlaufs dieser Vorlesung weiterentwickeln werden. Mein Ziel ist es, Ihnen die Gelegenheit zu bieten, einige dieser Tools während der Vorlesung, in der Nachbereitung oder Vorbereitung selbst auszuprobieren.

# **13 Vorkenntnisse und Erwartungen an die Studierenden**

## **13.1 Vertrautheit mit ChatGPT**

Ich bin mir sicher, dass ein Großteil von Ihnen bereits mit Tools wie ChatGPT vertraut ist oder sich eingehender damit beschäftigt hat. Der Begriff ChatGPT dürfte Ihnen kein Fremdwort sein und Sie wissen, wie man damit umgeht. In einer vorbereitenden Vorlesung im letzten Semester an der LMU München war ich erstaunt, als ich feststellte, dass bereits vor einem halben Jahr praktisch die gesamte Studierendenschaft mit ChatGPT vertraut war und es nutzte. Daher werden Sie in dieser Vorlesung keine Einführung in ChatGPT erwarten können, sondern ich setze diese Kenntnisse voraus. Stattdessen werde ich versuchen, tiefer in die philosophischen Aspekte der künstlichen Intelligenz einzutauchen.

## **13.2 Begriff der AI oder KI**

Bevor wir uns den Inhalten zuwenden, möchte ich Sie fragen: Was verstehen Sie unter AI? Lassen Sie uns kurz darüber nachdenken, bevor wir fortfahren.

# **14 Organisatorisches und Tools**

## **14.1 Vorlesungszeiten und -ort**

Die Vorlesung beginnt, obwohl in Agnes als 00 angekündigt, tatsächlich um CT. Diese Änderung ist nicht auf meinen Mist gewachsen, sondern geht auf die HU-Verwaltung zurück. Üblicherweise beginnt die Vorlesung um Viertel nach 10 und endet um Viertel vor 12, damit Sie ausreichend Zeit haben, zwischen den Veranstaltungen oder Universitäten zu wechseln.

## **14.2 Vorlesungswebseite und Materialien**

- In spätestens zwei Wochen werde ich eine eigene Webseite zur Vorlesung aufschalten. Ich verzichte auf die Nutzung von Moodle, da es für Lehrende ein Folterinstrument und eine Zeitverschwendungen darstellt.
- In der nächsten Woche werde ich Ihnen den Link zur Webseite hier zur Verfügung stellen.
- Auf der Webseite finden Sie für jede Vorlesung eine mit KI verfasste Zusammenfassung zum Herunterladen sowie weitere Materialien und gegebenenfalls Verlinkungen.

## **14.3 Zugriff auf Chat-GPT**

- OpenAI hat mir vorvergangene Woche mitgeteilt, dass sie die Zugriffe zu Chat-GPT freigeben. Eine Anmeldung sollte nicht mehr erforderlich sein.
- Ich konnte dies selbst nicht ausprobieren, da ich einen POE-Account besitze und eine Rückstufung zu kompliziert war.
- Ich würde mich über eine Rückmeldung von Ihnen freuen, ob der Zugriff bei Ihnen funktioniert.

## **14.4 Eigene KI-Webseite und virtueller Wittgenstein**

- In der zweiten Hälfte der Vorlesung werde ich möglicherweise eine eigene Webseite mit neuen KI-Möglichkeiten, die ich mit meinem eigenen Lab entwickle, freischalten und Ihnen zugänglich machen.

- Unter anderem planen wir, mit einem revitalisierten Wittgenstein mittels KI zu philosophieren.
- Ich hoffe, dass wir Ihnen einen virtuellen KI-Wittgenstein präsentieren können, mit dem Sie nicht nur Texte austauschen, sondern auch philosophische Diskussionen führen können.
- Weitere Details dazu werden wir in der zweiten Hälfte der Vorlesung erfahren.

## 14.5 Moodle und Teilnehmerliste

- Das Passwort für die Moodle-Vorlesung lautet “1234”. Dafür benötigen Sie keine künstliche Intelligenz, sondern lediglich ein gutes Gedächtnis.
- Bitte tragen Sie sich in die Teilnehmerliste ein.
- Im Falle von terminlichen Schwierigkeiten oder unvorhergesehenen Ereignissen wie Streiks, die mich am rechtzeitigen Erscheinen hindern, möchte ich Sie gerne per E-Mail erreichen können. Dies ist nur möglich, wenn Sie sich auf Moodle mit Ihrer E-Mail-Adresse registrieren.

## 14.6 Zulassung und Modulabschlussnoten

- ÜWP-Studierende, die durch unser automatisches Nicht-Intelligenz-System Agnes abgelehnt wurden, sollten sich nicht beunruhigen. Solange ausreichend Plätze vorhanden sind, lasse ich grundsätzlich alle Interessierten zu.
- Sollten Sie jedoch Modulabschlussnoten und Bescheinigungen benötigen, informieren Sie mich bitte im Vorfeld darüber, da es hierbei administrative Begrenzungen gibt, die Ihre Anwesenheit erfordern.

## **15 Nachfragen und individuelle Anliegen**

Ich stehe Ihnen am Ende der Vorlesung gerne für Nachfragen, Scheinanforderungen und ähnliche Anliegen zur Verfügung, solange ich nicht aus dem Vorlesungssaal geworfen werde. Um die wertvolle Vorlesungszeit optimal zu nutzen, möchte ich diese Themen nicht während der Veranstaltung behandeln. Individuelle Fragen beantworte ich auch gerne per E-Mail, wenn Sie mir schreiben. Liebe Studierende, zu Beginn der heutigen Vorlesung möchte ich ein paar organisatorische Dinge ansprechen. Wie Sie sich zu bestimmten Zeiten in Register eintragen und welche Erfordernisse für eine Prüfung als ÜWP-Studierende nötig sind, hängt von den verschiedenen Fakultäten ab. Bitte erkundigen Sie sich individuell darüber. Prinzipiell ist die Vorlesung natürlich offen und ich lasse alles zu, soweit es mir möglich ist. Wir müssen nur auf die administrativen Regelungen achten.

## 16 Was ist AI?

Künstliche Intelligenz, oder kurz AI, ist ein Begriff für eine technische Möglichkeit, die Mitte der 50er Jahre von einigen Kollegen aus Pittsburgh erdacht wurde. Ihr Ziel war es, maschinelle Computertechnologien zu entwickeln, die den menschlichen kognitiven Fähigkeiten nicht nur ebenbürtig sind, sondern sie sogar übertreffen. Man versprach damals vollmundig, dass dieses ehrgeizige Ziel in nur drei bis vier Jahren erreicht sein würde. Die Menschheit könnte dann endlich ihre Freizeit in vollen Zügen genießen, nur noch wenige Stunden pro Woche arbeiten, während der Rest von der KI erledigt würde.

Doch wie wir alle wissen, hat sich von dieser Vision bisher nichts eingelöst. Die Vorstellung war, dass KI als Meisterdisziplin des menschlichen Denkens schnell alle Bereiche überflügeln würde. Als Paradebeispiel galt damals das Schachspiel. Doch erst Anfang der 2000er Jahre gelang es einem Computerprogramm, den Schachweltmeister Garri Kasparov in einem ernsthaften Spiel zu besiegen - immerhin 50 Jahre später als ursprünglich prophezeit.

Das andere große Ziel, Computer zu entwickeln, die selbstständig wissenschaftlich kreativ denken können, ist bis heute nicht wirklich erreicht. Trotz aller anderslautenden, manchmal sensationsheischenden Meldungen bin ich jedoch sicher, dass diese Stufe in den nächsten Jahren erreicht werden wird. Dass also wissenschaftliche, kreative, kognitive und intellektuelle Aktivitäten von Maschinen alleine, ohne Assistenz von Forschern gemeistert werden. Das ist sozusagen noch die Krönung der Herausforderung von KI, von Artificial Intelligence.

## 17 KI als Verkaufsargument

Was Ihnen derzeit tagtäglich in der Öffentlichkeit als KI präsentiert wird, hat mit den eigentlichen Visionen und Zielen oft wenig zu tun. Nehmen wir als Beispiel eine Anzeige der Firma Samsung für ihre “Bespoke AI 11-Kilogramm-Washing-Maschine Serie 8 mit AI-Eco-Bubble und Quick-Drive”. Technisch gesehen handelt es sich schlicht um eine Waschmaschine, aber das Label “AI” soll den Verkauf ankurbeln.

Was ist daran nun wirklich AI? Nicht viel, es ist mehr ein Verkaufsargument als alles andere. Alles, was halbwegs gesteuert ist, wird heutzutage als AI vermarktet. Wenn ich hier “Licht aus” sage und es dunkel würde, würden Sie vielleicht denken “Oh, wir haben AI an der HU”. Dabei ist es letztlich nur eine etwas anspruchsvollere Steuerungstechnik, mehr nicht. Das Wort AI ist hier fehl am Platz, auch wenn es gerade en vogue ist.

## **18 Der Durchbruch der KI-Visionen**

Sind wir also jetzt in einer Zeit angekommen, in der sich die ursprünglichen KI-Visionen doch noch erfüllen könnten? Meine Antwort lautet: Ja. Und ich möchte Ihnen heute einen systematischen Grund dafür nennen, der für mich entscheidend ist und den ich Ihnen so vermitteln möchte, dass er nachvollziehbar wird. Nebenbei bemerkt: Wenn Sie Fragen oder Zwischenfragen haben, melden Sie sich einfach. Dann gestalten wir die Vorlesung etwas lebendiger und interaktiver.

Der Aspekt, auf den ich hinaus möchte und den ich für den Meilenstein halte, ist, dass die KI-Visionen gerade dabei sind Wirklichkeit zu werden. Die KI-Propaganda hingegen, die sollten wir schnell beiseite legen. Das ist in erster Linie ein Verkaufsargument, das nicht den Kern der technologischen Innovation ausmacht. Und genau das soll heute unser Thema sein.

# 19 Die Attraktivität von KI

Wo liegt denn potenziell die Attraktivität der KI, wie immer wir uns ihr auch nähern? Ist es eine bessere Internetsuchmaschine, die derzeit vielleicht eine der Triebfedern ist? Um das zu verstehen, müssen wir uns die Entwicklung des Internets vor Augen führen.

Gemessen an der Technologiegeschichte ist das Internet noch gar nicht so alt, etwas mehr als 20 Jahre. Wer die Anfänge noch miterlebt hat, erinnert sich an die ersten Browser, die damals oft mit Duschanlagen verwechselt wurden. Vor 20 Jahren wussten die wenigsten, was ein Internetbrowser eigentlich ist. Mittlerweile können wir uns ein Leben ohne Internet kaum noch vorstellen, weder technisch noch gesellschaftlich.

## 19.1 Die ursprüngliche Idee des Internets

Im Kern war die Konstruktion des Internets, die am CERN entwickelt wurde, folgende: Irgendwo stellen wissenschaftliche Einrichtungen webzugängliche Seiten als Informationsquellen bereit. Als Wissenschaftler oder technologische Provider verantworten sie die Inhalte, pflegen sie und sorgen für dauerhafte Zugänglichkeit. Die Browser sind lediglich das lesende Frontend für diejenigen, die auf die Inhalte zugreifen wollen.

Damals war das Internet also eine Art anspruchsvolles Faxgerät als Empfänger der Inhalte. Der Clou lag darin, dass man ganz einfach andere Inhalte per Verlinkung einbinden konnte. So entwickelte sich ein Schneeballsystem, das ein globales Netz von miteinander verknüpften Inhalten erzeugte. Das war die Webrevolution vor 20 Jahren.

## 19.2 Die Ablösung der Webwelt durch KI

Was wir jetzt erleben, ist eine Ablösung dieser Webwelt durch KI. In den nächsten Monaten werden Sie zunehmend feststellen, dass nicht mehr die Provider die Netzinhalte erstellen, auf Webservern bereitstellen und per Browser zugänglich machen. Diese Grundarchitektur wird abgelöst. Nicht mehr der Browser verantwortet, pflegt und stellt die Inhalte bereit. Das ist eine revolutionäre Änderung der Architektur der Informationsflüsse, aber auch der damit verbundenen Probleme. Einen Teil davon werden wir noch kennenlernen oder haben Sie schon erfahren.

Das Web funktionierte bisher deshalb, weil die Inhalte von den jeweiligen Personen, Institutionen oder Wissenschaftlern, die sie bereitstellten, auch autorisiert wurden. Für die Korrektheit und Richtigkeit bürgten die Glaubwürdigkeit und Gewissenhaftigkeit der Provider. Das ändert sich jetzt. Und wir alle wissen um die Gefahren, aber auch Potenziale, die damit einhergehen.

- Auf der einen Seite sind es nun große Internetfirmen, die die Inhalte über KI-Maschinen, sogenannte Bots, bereitstellen.
- Auf der anderen Seite können es auch böswillige Gestalten, Institutionen oder Staaten sein, die Inhalte generieren, ins Netz einspeisen, ohne als autorisierende Internetprovider in Erscheinung zu treten.

Derzeit wird das unter dem Stichwort “Internetinhalte der Social Media” diskutiert. Doch das ist nur die Oberfläche. Der Kern des Wandels und des Problems liegt darin, dass die Grundarchitektur des Internets mit den verantwortlichen Providern abgelöst wird durch - ich will nicht sagen unverantwortliche Bots - aber zumindest durch nicht mehr verantwortliche Internetinhaltsprovider. Und das hängt eben mit der KI-Revolution und dem Wandel der Informationsflüsse im Internet zusammen.<sup>#</sup> Die Veränderung der Informationssuche im Zeitalter der Künstlichen Intelligenz

Meine sehr verehrten Damen und Herren, lassen Sie uns heute gemeinsam einen Blick in die Zukunft der Informationssuche werfen. Bislang war es für uns alle selbstverständlich, dass wir bei der Suche nach Informationen auf die Dienste von Suchmaschinen wie Google zurückgreifen konnten. Wir vertrauten darauf, dass die von diesen autoritativen Anbietern bereitgestellten Inhalte glaubwürdig und sorgsam kuratiert waren. Doch in der nächsten Phase der digitalen Revolution wird sich dies grundlegend ändern.

### **19.3 Die Umgestaltung der Architektur des Internets**

Die Architektur des Internets befindet sich in einem extrem dynamischen Wandlungsprozess, dessen Ausgang noch niemand vorhersehen kann. Eines ist jedoch sicher: Es werden enorme Anstrengungen unternommen und gewaltige finanzielle Mittel investiert, um diese Transformation voranzutreiben. Jeder Staat, jede Region und auch Europa sollte ein vitales Interesse daran haben, die Kontrolle über diese Entwicklung nicht zu verlieren.

### **19.4 Neue Möglichkeiten durch Künstliche Intelligenz**

Doch lassen Sie uns zunächst einen Blick auf die vielversprechenden Möglichkeiten werfen, die uns die Künstliche Intelligenz eröffnet. Vielleicht erscheinen Ihnen einige dieser Anwendungen auf den ersten Blick trivial, doch ich versichere Ihnen, sie haben das Potenzial, unseren Alltag und unsere Arbeit grundlegend zu verändern.

#### **19.4.1 Hochwertige Übersetzungen**

Nehmen wir zum Beispiel das Thema Übersetzungen. Seit Jahrzehnten wurden enorme Ressourcen in die Entwicklung von linguistischen Modellen zur automatischen Übersetzung von Sprachen investiert. Doch lange Zeit waren die Ergebnisse bestenfalls als Partygags zu gebrauchen und keinesfalls für den ernsthaften Einsatz geeignet. In den letzten Jahren hat sich dies jedoch grundlegend geändert. Mittlerweile sind die automatischen Übersetzungen von so hoher Qualität, dass sie sogar für akademische Zwecke genutzt werden können.

Lassen Sie mich Ihnen ein Beispiel aus meinem eigenen Fachgebiet, der Wissenschaftsgeschichte, geben. Viele der historischen Quellen, mit denen wir arbeiten, sind in Latein verfasst. Vor 100 Jahren mussten Doktoranden ihre Dissertationen an unserer Fakultät noch auf Latein einreichen. Heute würden die meisten von Ihnen wohl Schwierigkeiten haben, einen lateinischen Quelltext sinnvoll zu interpretieren. Doch dank der Fortschritte in der Künstlichen Intelligenz gibt es Hoffnung. Vielleicht führen wir ja in unserer Fakultät bald wieder die Pflicht ein, Doktorarbeiten auf Latein zu verfassen - mit KI als Hilfsmittel könnte dies durchaus ein Alleinstellungsmerkmal unserer Universität werden.

#### **19.4.2 Simultanübersetzung und Lektoratsassistentz**

Die Möglichkeiten gehen jedoch noch weiter. In naher Zukunft werden wir in der Lage sein, hervorragende Simultanübersetzungen anzubieten. Ausländische Studierende, die keine europäische Sprache beherrschen, könnten meine Vorlesung mit einem Ohrhörer verfolgen und eine simultane Übersetzung erhalten.

Auch im Bereich des Lektorats gibt es spannende Entwicklungen. Programme wie Grammarly oder DeepL Write bieten bereits heute Textverbesserungsvorschläge, die durchaus mit der Qualität professioneller Lektoratsassistenzen mithalten können. Selbst große wissenschaftliche Verlage wie Nature stellen ihren Autoren mittlerweile Tools zur Verfügung, um ihre englischen Texte in lesbare Form zu bringen. Ob und wie dies gewünscht ist, wird derzeit heiß diskutiert. Doch ich bin davon überzeugt, dass in Zukunft das KI-gestützte Lektorat für wissenschaftliche Publikationen zum Standard werden wird.

#### **19.4.3 Automatisierte Forschungsberichte**

Vor der Tür stehen bereits Modelle, die in der Lage sind, eigenständig Texte wie Forschungsberichte zu verfassen. In experimentellen Wissenschaften wie der klinischen Forschung wird bereits daran gearbeitet, Ergebnisse und Erkenntnisse automatisch in Berichte zu überführen, die qualitativ den gängigen Publikationen entsprechen. Dies wirft natürlich Fragen auf:

- Wer ist der Autor eines solchen Berichts?
- Akzeptieren wissenschaftliche Journals Texte, die von einer KI erstellt wurden?

- Wie gehen wir mit Verantwortlichkeit, Seriosität und Zurechenbarkeit um?

Diese Probleme müssen gelöst werden, wenn wir diese Entwicklung weiter vorantreiben wollen.

## 19.5 Das Labor für gebildete KI

In meinem eigenen Labor, Lettre AI, setzen wir die Techniken um, die ich Ihnen in dieser Vorlesung vermitteln möchte. Unser Ziel ist es, eine KI bereitzustellen, die über die Fähigkeiten des Lesens, Übersetzens und Formulierens hinaus auch epistemische Qualifikationen mitbringt - also wissenbezogene Fähigkeiten, die wir gleich noch näher kennenlernen werden.

Lassen Sie mich Ihnen ein Beispiel für die bereits existierende Leistungsfähigkeit von KI geben. Ich zeige Ihnen hier einen Ausschnitt aus einem Werk, das zu Beginn des 17. Jahrhunderts wie ein Wirbelwind durch Europa fegte: den “Sidereus Nuncius” von niemand geringerem als Galileo Galilei. Dieses Buch markierte den Beginn einer Revolution, denn es war eines der ersten wissenschaftlichen Werke, das nicht nur auf Latein, sondern auch in der Volkssprache Italienisch verfasst wurde und so einer breiteren Öffentlichkeit zugänglich war.

Ich habe jetzt eine Variante von Chat-GPT aufgebaut. Für diejenigen unter Ihnen, die bereits mit Chat-GPT gearbeitet haben, wird die Oberfläche vertraut aussehen.

## 19.6 Der Kern der Vorlesung

Doch lassen Sie uns zum Kern dieser Vorlesung kommen. Mir ist es wichtig, dass wir gemeinsam verstehen, was KI eigentlich ist und wie sie funktioniert. # Begrüßung und Einführung in Interaktion mit ChatGPT

Stellen Sie sich vor, Sie öffnen die Seite von ChatGPT und werden mit der freundlichen, typisch amerikanischen Begrüßungsfloskel “How can I help you?” empfangen. Es klingt wesentlich servicefreundlicher als ein schlichtes “Hi, hier bin ich”. Die KI bietet Ihnen direkt ihre Hilfe an und präsentiert vier mögliche Optionen, die zwar oft irrelevant sind, Ihnen aber die Mühe ersparen sollen, sich selbst etwas auszudenken. Darunter können Sie dann eingeben, wobei Sie Unterstützung benötigen.

# **20 Demonstrieren der Möglichkeiten von ChatGPT anhand eines Beispiels**

## **20.1 Übertragen eines Bildes in maschinenlesbaren Text**

Nehmen wir an, Sie haben eine Seite mit komplexen Inhalten vor sich, mit denen Sie in ihrer jetzigen Form nichts anfangen können. Hier kommt die KI ins Spiel: Sie können einfach einen Screenshot der Seite machen und diesen in den Chat-GPT hochladen. Anschließend instruieren Sie die KI mit einer Anweisung wie “Transkribiere das Bild” - und schon erhalten Sie eine nahezu fehlerfreie Übertragung des nicht gerade einfachen Textes in getippte Buchstaben. Eine Leistung, die bis heute kein anderes Programm in dieser Qualität vollbringen kann.

## **20.2 Übersetzen des Textes in eine andere Sprache**

Doch das ist erst der Anfang. Nehmen wir an, Sie verstehen kein Latein - kein Problem. Tippen Sie einfach “Übersetze diesen Text ins Deutsche” ein und schon erhalten Sie eine verständliche, wenn auch noch etwas gewöhnungsbedürftige Übersetzung. Mit ein wenig Feinschliff oder dem Wechsel des Modells lässt sich daraus ein publikationsreifer deutscher Text erstellen. Und das Ganze funktioniert nicht nur für Deutsch und Englisch, sondern für über 150 Sprachen weltweit, darunter auch Japanisch und Koreanisch. Selbst obskure mittelalterliche Quellen stellen kein Hindernis dar.

# **21 Erweiterung der Möglichkeiten durch Phantasie und gezielte Fragestellungen**

Doch jetzt fängt der eigentliche Spaß erst an. Mit dem nun zugänglichen Text eröffnen sich ganz neue Möglichkeiten jenseits der typischen Google-Fragen wie "Wer war Galilei?" oder "Wann lebte er?". Stattdessen können Sie die KI mit Fragen herausfordern, die Google unmöglich beantworten kann. Zum Beispiel: "In welcher Stadt trank Galilei im Mai 1615 ein Glas Wein?". Das Problem liegt hier nicht nur darin, dass Google dieses spezifische Ereignis nicht kennt, sondern dass eine einfache Stichwortsuche prinzipiell nicht ausreicht, um die Antwort zu finden.

## **21.1 Analogie zu Sherlock Holmes**

Stellen Sie sich die KI als eine Art elektronischen Sherlock Holmes vor. Sie nimmt das gesamte Universum an Dokumenten über Galilei zur Kenntnis - seine Briefe, seine historischen Lebensumstände, seine typischen Aktivitäten im Frühjahr 1605. Aus diesen Informationen zieht sie dann Rückschlüsse und generiert eine fundierte Hypothese darüber, wo und wann Galilei wahrscheinlich sein Glas Wein genossen hat. Zwar nicht mit absoluter Sicherheit, aber basierend auf seinen regelmäßigen Lebensumständen. Solche Fragen werden die KI-Modelle in naher Zukunft beantworten können.

## **21.2 Vielfältige Analysemöglichkeiten von Texten**

Doch damit nicht genug. Sie können die KI auch anweisen, eine Tabelle mit allen Verben des Textes zu erstellen oder gezielt nach Verben zu suchen, die ein Lob, eine Ankündigung oder ein Versprechen ausdrücken - selbst wenn Sie die genaue Formulierung nicht kennen. Die Möglichkeiten sind schier grenzenlos.

Ein konkretes Beispiel: Fragen wir die KI, wer sich laut dem Text bewegt. Nach kurzer Bedenkzeit liefert sie die korrekte Antwort: Die vier Planeten bewegen sich zu verschiedenen Zeiten und mit erstaunlicher Geschwindigkeit um den Stern Jupiter - eine Entdeckung, die Galilei machte und die tatsächlich im lateinischen Originaltext erwähnt wird.

## 22 Philosophie als Grundlage für die Möglichkeiten der KI

Doch wie ist das alles möglich? Die Antwort liegt in der Philosophie - nicht in der Technik. Natürlich brauchen wir auch die technische Infrastruktur, so wie wir Beamer und Notebooks benötigen. Aber der eigentliche Schlüssel zu den Fähigkeiten der KI ist philosophischer Natur. Das wird oft übersehen, doch ich möchte Ihnen zeigen, warum Philosophie hier so entscheidend ist.

### 22.1 Beantwortung von Fragen über Mikrofoneingabe

Um das Potenzial der KI weiter zu verdeutlichen, können wir auch das Mikrofon aktivieren und eine Frage stellen: "Hat Galilei diese Entdeckung selbst durch Beobachtungen gemacht?". Das System denkt kurz nach und liefert dann die zutreffende Antwort: Ja, laut den Angaben im Text hat Galilei die Entdeckung tatsächlich selbst durch Beobachtungen gemacht.

Das Erstaunliche daran ist nicht nur, dass überhaupt eine Antwort generiert wird, sondern vor allem die Qualität dieser Antwort - trotz Versprechern und spontaner Formulierung mein-erseits.# Einführung in die sprachliche Dimension der KI

Meine Damen und Herren, heute möchte ich Ihnen eine faszinierende und zugleich beunruhigende Entwicklung in der Welt der künstlichen Intelligenz näherbringen. Es geht um die Fähigkeit von KI-Systemen, nicht nur Informationen aus autoritativen Quellen zu sammeln, sondern eigenständig Antworten zu generieren und Inhalte zu erstellen. Diese Entwicklung hat weitreichende Konsequenzen für unser Verständnis von Wissen und Informationsverarbeitung.

### 22.2 Die Möglichkeiten der KI

Die Möglichkeiten der KI sind atemberaubend und erweitern sich täglich. Lassen Sie mich Ihnen einige Beispiele nennen:

- Übersetzung: KI-Systeme können Texte von einer Sprache in eine andere übersetzen, und zwar mit einer Genauigkeit und Geschwindigkeit, die menschliche Übersetzer in den Schatten stellt.

- Bild-zu-Text-Konvertierung: KI kann Bilder analysieren und deren Inhalt in Textform beschreiben. Dies eröffnet völlig neue Möglichkeiten der Bildverarbeitung und -archivierung.
- Audio-zu-Text-Konvertierung: Gesprochene Sprache kann von KI-Systemen in Echtzeit transkribiert werden, was die Erstellung von Protokollen und Untertiteln erleichtert.
- Textzusammenfassung: Geben Sie der KI ein ganzes Buch, und sie wird Ihnen eine prägnante Zusammenfassung liefern. Dies kann die Recherche und das Studium enorm beschleunigen.
- Text-zu-Audio-Konvertierung: Umgekehrt kann KI auch geschriebenen Text in gesprochene Sprache umwandeln, was neue Möglichkeiten für Hörbücher und Sprachassistenten eröffnet.
- Text-zu-Video-Konvertierung: Hier wird es geradezu unheimlich. KI kann aus Textbeschreibungen realistische Videos generieren, die kaum noch von echten Aufnahmen zu unterscheiden sind.

## 22.3 Die Gefahren der KI

So faszinierend diese Möglichkeiten auch sind, sie bergen auch erhebliche Risiken. Ein zentrales Problem ist das Phänomen der “Halluzination”. Dabei generiert die KI scheinbar plausible Informationen, die jedoch nicht der Realität entsprechen.

Ein Beispiel: Ich fragte eine KI nach dem Namen der zweiten Frau des Mathematikers Leonhard Euler. Die Antwort klang überzeugend, inklusive eines Verweises auf eine Publikation der Petersburger Akademieschriften von 1784. Doch diese Publikation existiert gar nicht, und die genannte Person war nie mit Euler verheiratet.

Solche Halluzinationen können fatale Folgen haben, wenn sie unerkannt bleiben. Wer eine solche Information zitiert, disqualifiziert sich wissenschaftlich für immer. Dieses Problem trat auch bei der Mars-Mission der NASA auf, als eine KI falsche Informationen über einen Erkundungssatelliten verbreitete.

## 22.4 Der sprachliche Kern der KI

Bei all diesen Anwendungen, sei es Bild-, Audio- oder Videoverarbeitung, bildet die Sprache den Kern der KI-Technologie. Selbst bei der Bildanalyse übersetzt die KI zunächst das Bild in eine verbale Beschreibung, bevor sie weiterverarbeitet wird.

Diese Erkenntnis ist philosophisch bedeutsam und erinnert an Wittgensteins These von der Unhintergehrbarkeit der Sprache. Die sprachliche Verbalisierung von Inhalten ist der Dreh- und Angelpunkt der KI, und genau darum soll es in dieser Vorlesung gehen.

Ich werde mich nicht auf die technischen Details der KI-Entwicklung konzentrieren, sondern auf den Umgang mit Sprache in KI-Modellen. Die anderen Medien sind zwar faszinierend, aber letztlich sekundär. Unser roter Faden wird die philosophische Dimension der sprachlichen Verarbeitung in der KI sein.<sup>#</sup> Gefahren und Probleme der künstlichen Intelligenz

Meine Damen und Herren, lassen Sie uns heute über die Schattenseiten der künstlichen Intelligenz sprechen. Wir haben bereits die atemberaubenden Möglichkeiten dieser Technologie gesehen, doch nun ist es an der Zeit, auch die Probleme und Gefahren zu beleuchten, die damit einhergehen.

## 22.5 Das Problem der Halluzinationen

Eines der ersten Probleme, auf das wir stoßen, sind die sogenannten Halluzinationen der KI-Modelle. Ein eindrucksvolles Beispiel dafür lieferte das Supermodell von Google, das auf die Frage "Wer fliegt denn da?" eine Antwort gab, die zwar plausibel klang, aber rein fiktiv war. Ohne Zugriff auf aktuelle NASA-Informationen oder Tagesnachrichten erfand das Modell kurzerhand einen Satellitennamen. Innerhalb einer halben Stunde wurde es vom Netz genommen, und der Marktwert von Google-Aktien sank um Millionen. Seitdem trauen sich die Unternehmen nicht mehr, ihre Modelle zu veröffentlichen.

Doch warum halluzinieren die Modelle überhaupt, wenn sie doch schon so viele Fähigkeiten besitzen? Die Antwort darauf ist komplexer als man denkt.

## 22.6 Die Gefahr der Manipulation durch glaubwürdige Fakes

Ein weiteres Problem, das eng mit den Halluzinationen verbunden ist, ist die Fähigkeit der KI, glaubwürdige Texte, Bilder und sogar Videos zu produzieren. Dies öffnet Tür und Tor für falsche oder manipulative Informationen, die auf den ersten Blick echt erscheinen.

Ein aktuelles Beispiel dafür sind die Videos, die im Zusammenhang mit dem Raketenüberfall auf Israel in den sozialen Medien aufgetaucht sind. Sie zeigten panische Einwohner von Tel Aviv, die vor nicht existierenden Einschlägen flohen. Diese Videos wurden absichtlich generiert, um die Öffentlichkeit zu täuschen, und sind für den Betrachter zunächst nicht als Manipulation zu erkennen.

## **22.7 Selektive Informationen und die Pluralität der Hintergründe**

Jede Antwort, die uns ein KI-Modell gibt, basiert auf bestimmten Annahmen und Voraussetzungen. Diese haben jedoch immer auch Alternativen, die möglicherweise nicht besser oder schlechter sind, aber eine Pluralität an Hintergründen darstellen.

Wenn wir eine bestimmte Antwort akzeptieren, akzeptieren wir auch die Voraussetzungen dafür und vernachlässigen die Alternativen. Ein Beispiel dafür ist die Anfrage an ein KI-Modell, ein Porträt eines möglichen Nachfolgers des jetzigen Papstes zu erstellen. Aufgrund der politisch korrekten Voreinstellung des Modells wurde eine farbige Frau im Papstgewand generiert - eine Darstellung, die in der Realität aufgrund der Zusammensetzung des Kardinalskollegiums höchst unwahrscheinlich ist.

Dieses Beispiel verdeutlicht, wie selektive Informationen zu verzerrten Ergebnissen führen können. Es wirft die Frage auf, wie wir mit diesen Problemen umgehen sollen.

## **22.8 Die Unausweichlichkeit der KI-Entwicklung und die Notwendigkeit der Gestaltung**

Eines ist klar: Wir können uns vor diesen Fragen nicht drücken. Die Entwicklung der künstlichen Intelligenz ist unwiderstehlich und unausweichlich. Ab heute werden uns diese Technologien mit all ihren Vor- und Nachteilen zunehmend beschäftigen.

Wir müssen lernen, damit umzugehen und die Entwicklung aktiv mitzugestalten. Nicht im Sinne einer Kontrolle, sondern einer Gestaltung. Denn wenn wir jetzt nicht eingreifen, laufen wir Gefahr, die Kontrolle über diesen Prozess zu verlieren.

## **22.9 Weitere Gefahren: Diskriminierung und Überwachung**

Neben der selektiven Information gibt es weitere Gefahren, die wir im Auge behalten müssen. Dazu gehören Dimensionen der Diskriminierung, bei denen bestimmte Personengruppen oder Qualifikationen berücksichtigt werden, andere hingegen nicht.

Auch die Möglichkeiten der Überwachung durch KI-Systeme sind alarmierend. Ein Beispiel dafür ist China, wo Besucher bei der Einreise lediglich in eine Kamera lächeln müssen und dann während ihres gesamten Aufenthalts live verfolgt und protokolliert werden.

Diese Entwicklungen werfen Fragen auf, wie weit solche Technologien zugelassen und kontrolliert werden sollten. Eine Antwort darauf zu finden, ist keine leichte Aufgabe.

## 22.10 Die Notwendigkeit der Auseinandersetzung mit KI

Angesichts dieser erschütternden Probleme könnte man geneigt sein, das Thema KI einfach zu vergessen. Wozu sich mit Übersetzungen von Galileis lateinischen Texten beschäftigen, wenn wir dafür doch unsere Gelehrten haben?

Doch so einfach ist es nicht. Die Vorteile der künstlichen Intelligenz sind zu groß, um sie zu ignorieren. Wir müssen uns mit dieser Technologie auseinandersetzen, ihre Möglichkeiten nutzen und gleichzeitig ihre Schattenseiten im Blick behalten. Nur so können wir eine Zukunft gestalten, in der die KI zum Wohle der Menschheit eingesetzt wird. # Begrüßung und Einführung

Einen schönen guten Tag, meine Damen und Herren. Heute möchte ich mit Ihnen über zwei Fragen sprechen, die mir in letzter Zeit immer wieder begegnen. Zunächst einmal habe ich eine Frage zu der Konferenz, von der ich gehört habe, dass sie in diesem Monat August stattfinden soll. Wo genau findet diese Konferenz statt?

Ah, ich verstehe. Es handelt sich also um eine regelmäßig wiederkehrende Konferenzserie, die im August abgehalten wird. Es ist bemerkenswert, wie weit fortgeschritten die Aufmerksamkeit und das Wissen um diese Themen inzwischen auch in den Institutionen sind. Sogar auf EU-Ebene wurde im März letzten Jahres bereits ein Bericht veröffentlicht, in dem diese Angelegenheiten thematisiert wurden. Allerdings wurden sie dort so behandelt, wie ich es gerade geschildert habe - sie wurden angesprochen, aber nicht gelöst.

Es gibt keine einzelne Konferenz, die sich des Problems annimmt und von der wir erwarten könnten, dass es in kürzester Zeit gelöst wird. Nein, so einfach ist es leider nicht. Stattdessen möchte ich auf die philosophischen Aspekte eingehen, die sowohl den Vorteilen als auch den Gefahren zugrunde liegen.

## 23 Beispiele für die Nutzung von Sprachen in der Wissenschaft

Lassen Sie uns ein Beispiel betrachten, das wir gerade schon diskutiert haben. In der Fachliteratur hält sich hartnäckig das Gerücht, dass Galileis Vater sich negativ über die wissenschaftliche Nutzung anderer Sprachen als Latein geäußert haben soll. Das würde natürlich einen spannenden Vater-Sohn-Konflikt darstellen, denn Galilei selbst ist ja berühmt dafür, dass er das Italienische für die Wissenschaft nutzbar machte, indem er auf Italienisch publizierte.

In zahlreichen Sekundärquellen findet man die These, dass sein Vater dies nicht für wissenschaftlich hielt und dass sein Sohn Galileo Galilei sich besser von diesen italienischen Publikationen fernhalten sollte. Oh, Moment mal - da steht, dass Kepler sich gegenüber Galilei negativ geäußert hat, nicht Galileis Vater. Danke für den Hinweis! Das ist keine Halluzination, sondern ein echter Fehler meinerseits. Ich hoffe, ich vergesse nicht, das für die Internetversion zu korrigieren.

Die Pointe ist jedenfalls, dass man eine solche Frage - ob sich eine Person X irgendwo negativ zu einer bestimmten These geäußert hat - mit Google nicht beantworten kann. Das mag trivial klingen, aber im Moment ist es tatsächlich nicht möglich, dies durch eine Google-Suche herauszufinden. Warum? Weil Google Ihnen kein Dokument im Internet liefern wird, in dem diese Frage direkt beantwortet wird. Und wenn es ein solches Dokument nicht gibt, ist die Frage für Sie mit Google-Techniken nicht zu beantworten.

Dabei handelt es sich um eine Frage, die historisch gesehen entweder wahr oder falsch ist. Wie kann man das also entscheiden? Nicht mit den heutigen Google-Techniken. Hier braucht es eine neue Dimension der Recherche, die über bestimmte Fähigkeiten verfügen muss.

## **24 Aufgaben und Fragen, die mit herkömmlichen Methoden nicht lösbar sind**

Lassen Sie mich Ihnen anhand einer Liste von Aufgaben und Fragen veranschaulichen, wie zunehmend Probleme auftauchen, die mit den heutigen akademischen Techniken nicht zu lösen sind. Ich spreche hier von Fragen, die selbst Sie als forschende Person nicht beantworten können, wenn sie halbwegs komplex sind.

Mir geht es um die unlösbaren Probleme der realen Forschungswelt, die zwar mit KI lösbar wären, aber aufgrund bestimmter fehlender Fertigkeiten bisher nicht gelöst werden können. Jetzt befinden wir uns im philosophischen Teil meiner Ausführungen und ich werde versuchen, dies sprachanalytisch zu komprimieren.

### **24.1 Frage 1: Einfache Aussage in einer Quelle**

Angenommen, Person A äußert sich in einer Quelle Q zu einer Person namens Jochen Schmidt. Ist diese Aussage wahr oder falsch? Hier haben Sie noch eine gewisse Chance, die Frage eindeutig zu beantworten, wenn Sie die Quelle Q gefunden haben und darin die Person A benannt wird und sich zu Jochen Schmidt äußert. Der Anforderungsgrad ist hier noch nicht sehr hoch. Wenn das Ihre Examensaufgabe wäre, hätten Sie eine realistische Chance, sie zu lösen. Sie müssten nur so lange alle Quellen durchlesen, bis Sie die richtige gefunden haben.

### **24.2 Frage 2: Aussage in Briefen zu einem Thema**

Nehmen wir an, Person A äußert sich in ihren Briefen zu einem Thema T. Das können Sie schon nicht mehr ohne weiteres lösen, ohne eine Lebensdauer damit zu verbringen, das gesamte Schrifttum von Person A zu lesen. Wenn Sie z.B. für eine Examsarbeit eine Biografie über eine Person namens Heinz Müller verfassen sollten und eine solche Aufgabe hätten, müssten Sie zunächst alle Briefe zusammentragen und sie komplett lesen. Und selbst dann wären Sie sich nicht sicher, ob Sie wirklich alle Briefe gefunden haben.

Denken Sie nur an die Kafka-Forscher. Wenn Sie wissen wollen, ob sich Kafka in seinen Briefen jemals zu einem bestimmten Thema geäußert hat oder nicht, haben Sie einen enormen manuellen Forschungsaufwand vor sich, um überhaupt in die Nähe einer Antwort zu kommen.

Hier befinden wir uns bereits in Bereichen, die schwer zu beantworten sind - Fragestellungen, die bislang praktisch nicht zu lösen waren.

### **24.3 Frage 3: Aussagen einer Person in ihren Schriften**

Hat eine Person A in ihren Schriften Aussagen der Art T getroffen, wenn Person A sehr viel geschrieben hat? Nehmen wir als Beispiel die Briefe Napoleons. Hat sich Napoleon jemals zu Aspekten der Vorläufer der Genfer Konvention bei der Kriegsführung geäußert? Das können Sie aus praktischen Gründen nicht lösen. Ich will an dieser Stelle nicht sagen, dass es prinzipiell unmöglich ist, aber in der Wissenschaft möchte man solche Fragen beantwortet haben. Und das gilt nicht nur für das öffentliche Interesse, sondern auch für die Wissenschaft selbst.

Sie können sich vorstellen, Welch enorme Konsequenzen es für die Wissenschaft hätte, wenn man solche Fragen überhaupt beantworten könnte. Dann wäre es möglich, weitreichende Thesen zu Napoleons Verständnis von Krieg und Frieden aufzustellen, die von der Evidenz abhängen, mit der man solche Fragen beantworten kann. Im Moment ist das nicht möglich.

### **24.4 Frage 4: Keine Aussage einer Person in ihren Schriften**

Angenommen, Person A hat in ihren Schriften keine Aussage T getroffen. Als normaler arbeitender Historiker oder Geisteswissenschaftler werden Sie diese Frage nicht seriös beantworten können. Deshalb gibt es in der Literatur die Unsitte, andere Werke zu zitieren, die sich aus irgendwelchen Gründen dazu bemüßt fühlten, solche Fragen zu beantworten.

Ein Beispiel: Nehmen wir wieder Kafka. Manche Autoren vertreten die These, dass Kafka sich nie antisemitisch geäußert hat. Aber welche Evidenz können Sie dafür eigentlich angeben? Es ist schwierig, eine nicht vorhandene Lektüre von Briefen als Beleg anzuführen. Wie wollen Sie eine solche These rechtfertigen, wenn Sie sie vertreten?

Eine der größten Unsitten der gegenwärtigen akademischen Literatur besteht darin, nicht selbst das Risiko einer These einzugehen, sondern stattdessen den berühmten Heinz Müller zu zitieren, weil er schon einmal etwas Ähnliches gesagt hat. Also fügt man eine Fußnote in die Arbeit ein: "Heinz Müller, 1973, Seite 5: Ganz klar, Kafka hat sich nie antisemitisch geäußert." Und auf einmal entsteht ein Schneeballsystem, das dem Halluzinationseffekt ähnelt, den wir gerade hier hatten. Und zwar nur deshalb, weil die Evidenz, die für bestimmte Thesen erforderlich ist, auf manuelle Weise kaum zu beschaffen ist. Mit KI werden Sie das in Zukunft können.

# **25 Die Herausforderung der inhaltlichen Analyse mit KI**

Jetzt werden Sie vielleicht fragen: Inwiefern ist das speziell für KI relevant? Man könnte doch erwarten, dass sich das grammatisch lösen lässt. Wenn ich die Aussage T formalisieren kann, müsste ich doch auf dem Textkorpus einfach prüfen können, ob diese Bedingung irgendwo erfüllt ist, oder?

Genau das ist der springende Punkt, und ich muss jetzt ein bisschen auf die Uhr schauen, damit ich meine Kurve hier noch hinbekomme. Aber diese Kurve berührt schon das Thema. Was heißt es, in Ihrem Korpus prüfen zu können?

Nehmen wir an, Sie hätten den Idealfall: Kafkas gesammelten Briefwechsel in einer Datenbank. Jetzt möchten Sie wissen, ob es darin eine antisemitische Formulierung gibt. Wie sieht die denn aus? Wenn Sie Ihre Datenbank nach Art einer Google-Suche nach bestimmten Wortvorkommnissen durchforsten, dann können Sie das lösen. Das ist die klassische Vorgehensweise.

Aber inhaltlich betrachtet: Was ist eigentlich eine antisemitische Äußerung? Sobald es darum geht - und deshalb habe ich es hier erwähnt - kön# Betrachtungen zur künstlichen Intelligenz und Sprachverarbeitung

Meine sehr geehrten Damen und Herren, liebe Studierende,

in der heutigen Vorlesung möchte ich Ihnen einen faszinierenden Einblick in die Welt der künstlichen Intelligenz und insbesondere deren Fähigkeiten zur Sprachverarbeitung geben. Wir werden uns mit der Frage beschäftigen, inwieweit KI-Systeme in der Lage sind, komplexe sprachliche Konstrukte wie Metaphern, Ironie oder versteckte Bedeutungen zu erkennen und zu interpretieren.

## **25.1 Grenzen der traditionellen Datenbanken**

Zunächst einmal möchte ich klarstellen, dass ich keineswegs behauptet habe, es gäbe in den vorliegenden Dokumenten keine relevanten Satzvorkommnisse. Die herkömmliche Art der Dokumentenaufzeichnung und -abfrage, wie sie etwa mit Datenbanken möglich ist, erlaubt zwar das Auffinden bestimmter Textpassagen, jedoch keine inhaltlichen Suchen im eigentlichen Sinne.

Selbst moderne KI-Systeme können nicht mit absoluter Sicherheit feststellen, dass eine bestimmte Aussage nicht getroffen wurde, da stets die Möglichkeit besteht, dass die zugrunde

liegende Datenbasis unvollständig ist. Vielmehr lässt sich hier nur mit Wahrscheinlichkeiten operieren - ein Begriff, den ich an dieser Stelle allerdings kritisch hinterfragen möchte.

## **25.2 Qualifizierte Aussagen auf Basis der verfügbaren Evidenz**

Wahrscheinlichkeiten sind numerische Werte zwischen 0 und 1, die man in diesem Kontext nicht sinnvoll einsetzen kann. Stattdessen sollte man sich auf die konkrete Situation beziehen und feststellen: Auf Basis dieser und jener Grundgesamtheit von Briefwechseln und Äußerungen, die als Dokumente für die Befunde zur Verfügung stehen, lässt sich unter der Voraussetzung, dass sie die alleinige Entscheidungsgrundlage bilden, folgendes Fazit ableiten.

Eine solche differenzierte Betrachtung der Befundlage ist unerlässlich, denn es lässt sich ja nicht ausschließen, dass genau jene Briefe, die möglicherweise relevante Inhalte enthalten, vernichtet wurden. Ein solches Szenario würde den Wahrheitswert der Fragestellung grundlegend verändern. Auch KI-Systeme können diese Problematik nicht vollständig ausräumen, sehr wohl aber eine qualifizierte, auf der verfügbaren Evidenz basierende Antwort geben.

## **25.3 Herausforderungen bei der Interpretation von Metaphern und Ironie**

Ein besonders spannendes Feld ist die Fähigkeit von KI-Systemen, mit Metaphern und ungewöhnlichem Sprachgebrauch umzugehen. Gerade im Kontext des Antisemitismus verbergen sich oft codierte Botschaften hinter scheinbar harmlosen Formulierungen. Während eine Blut- und Boden-Ideologie relativ leicht zu identifizieren ist, stellt die Interpretation von Begriffen wie "entwurzelt" oder "ohne Verwurzelung" eine ungleich größere Herausforderung dar.

Anhand eines konkreten Beispiels möchte ich Ihnen verdeutlichen, wozu moderne KI-Systeme in diesem Bereich bereits in der Lage sind. In München hatten wir es mit revolutionären Briefen aus der Zeit der Französischen Revolution zu tun, die in elegantem Französisch verfasst waren und vor Ironie und Sarkasmus nur so strotzten. Um diese Feinheiten zu erkennen, bedarf es zunächst einmal exzellenter Sprachkenntnisse. Doch selbst dann gilt es, die ironischen Komponenten als solche zu identifizieren.

Ich kann Ihnen versichern, dass KI-Systeme mittlerweile über eine Sprachkompetenz verfügen, die es ihnen erlaubt, auch diese Dimension der Sprachverwendung zu erkennen. Allerdings dürfen Sie sich das nicht als simples Schwarz-Weiß-Schema vorstellen, bei dem man einfach einen "Ironie-Kompetenz-Knopf" umlegt und schon funktioniert alles wie bei einem literarischen Meisterinterpreten.

## **25.4 Lernfähigkeit und Entwicklungspotenzial von KI-Systemen**

Vielmehr müssen Sie sich den Lernprozess der KI ähnlich vorstellen wie Ihre eigene Entwicklung zu Beginn Ihres Studiums. Auch Sie haben im Laufe der Zeit eine Menge dazugelernt und sich weiterentwickelt. Genauso können auch KI-Modelle lernen und sich verbessern. Ich möchte keineswegs behaupten, dass bereits alle Probleme und Herausforderungen gelöst sind, aber es gibt vielversprechende Lösungsansätze, um auch mit komplexeren Formen der Sprachverwendung umgehen zu können.

In München haben wir beispielsweise erfolgreich getestet, ob KI-Systeme in der Lage sind, bissige Karikaturen aus den 1920er Jahren zu interpretieren und zu erkennen, welche Personen mit welchen Klischees auf den Arm genommen werden. Mit dem richtigen Training ist es den Bilderkennungsalgorithmen tatsächlich gelungen, diese Zusammenhänge zu entschlüsseln.

## **25.5 Der Paradigmenwechsel durch Large Language Models und Embeddings**

Der entscheidende Unterschied und gleichzeitig der Punkt, an dem der “Philosophical Turn” der KI einsetzt, liegt in der Entwicklung von Techniken wie Large Language Models oder Embeddings. Diese ermöglichen eine Abkehr von der reinen Textsuche hin zu einer inhaltlichen Erfassung der Bedeutung sprachlicher Ausdrücke. Dieser semantische Wechsel, den ich auch gerne als “Semantic Turn” bezeichne, ist der Schlüssel zu den beeindruckenden Fähigkeiten moderner KI-Systeme.

Egal ob es um die Analyse von Bildern, Texten oder Audioaufnahmen geht - all diesen Anwendungen liegt zugrunde, dass die Systeme nicht nur nach bestimmten Zeichenfolgen suchen, sondern deren Bedeutung erfassen und identifizieren können. Genau darum geht es bei den milliardenschweren Investitionen in diesem Bereich: den Modellen beizubringen, auf Basis der eingegebenen Daten die dahinterstehende Semantik zu erkennen.

## **25.6 Die Bedeutung der Philosophie für die KI-Forschung**

Damit eröffnet sich ein weites Feld für die Philosophie. Solange wir nur von Sätzen sprechen, bewegen wir uns auf der Ebene von Formulierungen und syntaktischen Strukturen. Wenn wir jedoch nach der Bedeutung eines Ausdrucks fragen, betreten wir Neuland. Genau hier setzt die aktuelle KI-Revolution an, und deshalb ist die Philosophie von zentraler Bedeutung für diese Entwicklung.

Als Studierende der Philosophie sollten Sie mit der klassischen Unterscheidung zwischen Satz und Aussage vertraut sein. Im Deutschen ist diese Differenzierung von größter Wichtigkeit, während sie in englischen Übersetzungen oft vernachlässigt wird. So haben etwa die Übersetzer

von Wittgensteins Gesammelten Werken sowohl für “Aussage” als auch für “Satz” durchgängig den Begriff “Sentence” verwendet, was zu erheblichen Missverständnissen führen kann. Im Englischen heißt es korrekterweise “Sentence” für Satz und “Proposition” für Aussage.

Genau diese Unterscheidung markiert die fundamentale Revolution, die sich gerade vollzieht: Wir haben es nun mit Maschinen zu tun, die mit Aussagen umgehen können. Und nur Aussagen, nicht Sätze, können wahr oder falsch sein. Wer also über Fake News, Halluzinationen und ähnliche Phänomene spricht und sich dabei auf Sätze bezieht, liegt philosophisch gesehen völlig falsch. Wahrheit und Falschheit können sich konzeptionell nur auf Aussagen beziehen.

Die Tatsache, dass KI-Systeme nun in der Lage sind, sich mit Aussagen zu befassen, birgt ebenso faszinierende Möglichkeiten wie Gefahren. In der nächsten Vorlesung werden wir uns eingehender mit diesen Aspekten beschäftigen und uns ansehen, wie genau diese neuen Technologien funktionieren und welche Auswirkungen sie haben können.

Ich danke Ihnen für Ihre Aufmerksamkeit und freue mich darauf, dieses spannende Thema in der kommenden Woche gemeinsam mit Ihnen zu vertiefen. # Begrüßung zur zweiten Vorlesung Philosophie der AI

Herzlich willkommen, meine Damen und Herren, zur zweiten Vorlesung unserer Reihe “Philosophie der AI”. Lassen Sie uns heute an die spannenden Erkenntnisse der letzten Sitzung anknüpfen und gemeinsam ergründen, welche faszinierenden Möglichkeiten die Künstliche Intelligenz für die geisteswissenschaftliche Forschung bereithält. Stellen Sie sich vor, wie AI unsere alltägliche Arbeit nicht nur erleichtern, sondern revolutionieren und bisher ungeahnte Perspektiven eröffnen kann.

# **26 Traditionell schwer lösbarer Fragen in der Forschung**

In der Welt der Wissenschaft gibt es eine Vielzahl von Fragestellungen, die uns immer wieder vor Herausforderungen stellen und deren Beantwortung mit herkömmlichen Mitteln oft an Grenzen stößt. Nehmen wir beispielsweise die Suche nach Evidenz in einem definierten Kreis von Quellen, einem sogenannten Scholarium, um eine historische Aussage H zu belegen. Jeder von Ihnen, der schon einmal eine wissenschaftliche Arbeit verfasst hat, weiß, wie zeitaufwändig und mühsam dieser Prozess sein kann - je nach Komplexität der Fragestellung. Doch mit der Unterstützung von AI könnten wir in Zukunft, abhängig von der Zugänglichkeit und Aufbereitung des Scholariums, solche Nachweise schnell und effizient führen.

## **26.1 Evidenz finden, um eine Hypothese zu widerlegen**

Noch kniffliger wird es, wenn wir in einem Scholarium nach Evidenz suchen, um eine Hypothese H zu widerlegen. Im Alltag des Wissenschaftlers ist dies praktisch unmöglich - und dennoch finden wir solche Aussagen häufig in Publikationen. Überlegen Sie selbst: Wie oft haben Sie schon in Hausarbeiten, Qualifikationsschriften oder Fachartikeln Behauptungen gelesen, die eine These anhand von Standardreferenzliteratur zu widerlegen versuchen? Häufig sucht man vergeblich nach der tatsächlichen Evidenz dafür. Stattdessen wird allzu oft der bequeme Weg gewählt, sich auf Kollegen zu berufen, die ähnliche Aussagen getroffen haben - doch das ist keine echte Evidenz, sondern bestenfalls eine fragwürdige Praxis.

## **26.2 Zeitgenössische Autoren und ihre Äußerungen zu historischen Hypothesen**

Lassen Sie uns noch einen Schritt weiter gehen und uns einer noch komplexeren Fragestellung zuwenden: Welcher zeitgenössische Autor hat sich zu einer historischen Hypothese H ebenfalls geäußert? Stellen Sie sich vor, Sie interessieren sich für eine bestimmte These, die der wissenschaftshistorische Autor Johannes Kepler im Jahre 1603 aufgestellt hat. Nun möchten Sie wissen, welche seiner Zeitgenossen sich zu ähnlichen Fragen geäußert haben. Ohne jahrelange akribische Lektüre und Archivarbeit ist eine solche Recherche praktisch unmöglich.

## **26.3 Der Einfluss von Publikationen auf historische Autoren**

In wissenschaftlichen Feststellungen stoßen wir oft auf Aussagen wie: "Wer hat die Publikation von H, eines historischen Autors, relevant beeinflusst?" Doch Hand aufs Herz - die meisten dieser Behauptungen sind spekulativ und unbegründet. Nicht etwa, weil die Forscher unseriös arbeiten, sondern weil der Evidenznachweis für solche Aussagen extrem schwierig zu führen ist. Schon allein die Frage, was genau einen "relevanten Einfluss" auf eine Hypothese ausmacht, ist alles andere als trivial.

# **27 Die Rolle der AI in der geisteswissenschaftlichen Forschung**

Hier zeigt sich deutlich, dass die Diskussion um AI weit über eine rein technische Erleichterung unserer Arbeit hinausgeht. Vielmehr eröffnet sie uns die Möglichkeit, alltägliche Fragestellungen überhaupt erst bearbeitbar zu machen, die bislang nur unzureichend gelöst werden konnten. Nehmen wir als weiteres Beispiel die Frage, wer eine Alternative zu einer historischen Hypothese H vertreten hat. Schon die Definition dessen, was eine “Alternative” in diesem Kontext bedeutet, ist eine Herausforderung - von der Suche nach entsprechenden Äußerungen in der Gesamtliteratur eines Scholariums ganz zu schweigen. Praktisch unmöglich, wenn auch theoretisch denkbar.

## **27.1 Die Gefahren der AI und ihre Korrektur durch verbesserte Praktiken**

In der letzten Vorlesung haben wir unter der Rubrik “Gefahren der AI” diskutiert, wie Aussagen oder Befunde, die mittels AI generiert wurden, selektiv sein können, halluzinierte Thesen vertreten oder auf andere Weise kritisch hinterfragt werden müssen. Doch heute betrachten wir die Kehrseite der Medaille: AI kann auch dazu beitragen, unzulängliche oder problematische Praktiken in der gegenwärtigen Forschung zu korrigieren oder gar gänzlich zu ersetzen. Meine These lautet daher unmissverständlich: Der Eingriff von AI in unser wissenschaftliches Tagesgeschäft wird unsere Disziplinen in kürzester Zeit, in wenigen Jahren, drastisch verändern. Mein Rat an Sie: Beschäftigen Sie sich so schnell wie möglich mit diesen Mitteln, auch schon während Ihres Studiums - andernfalls werden viele Fragen Ihrer Qualifikationsarbeiten nicht mehr den zukünftigen Anforderungen genügen.

## **27.2 Nachvollziehbarkeit von Begründungen für historische Hypothesen**

Ein weiteres Beispiel für eine bislang schwer zu beantwortende Frage ist, inwiefern die Begründung für eine historische Hypothese H für andere Zeitgenossen nachvollziehbar oder überzeugend gewesen sein mag. Wenn wir wissenschaftliche Kontroversen einer bestimmten Epoche verstehen wollen, müssen wir uns fragen: Warum konnte die Publikation eines Autors A seine

Kollegen B nicht überzeugen? Ein klassischer Fall in der Wissenschaftsgeschichte ist das Werk “De revolutionibus” von Kopernikus, das zwar einige, aber bei weitem nicht die Mehrheit seiner Zeitgenossen überzeugen konnte. Doch warum war das so? Spekulationen führen uns hier nicht weiter - stattdessen müssen wir solche Fragen auf eine solide methodische Grundlage stellen, und dies ist nur mittels KI möglich.

# **28 Die Entwicklung der KI und ihr Einfluss auf das wissenschaftliche Arbeiten**

Doch wie genau kann die KI uns bei der Beantwortung solch komplexer Fragen unterstützen? Dieser Frage werden wir uns im Laufe der Vorlesung eingehend widmen - und ich bin zuverlässig, dass wir gemeinsam Antworten finden werden. Dabei werden wir feststellen, dass die Anwendung von KI weit weniger komplex und kompliziert ist, als man zunächst vermuten mag. Vielmehr ist sie das Ergebnis eines Jahrzehntelangen Entwicklungsprozesses, der nun zu einem Leistungssprung führt, der auf den ersten Blick wie eine einmalige technische Neuerung aus dem Nichts erscheinen mag. Doch dieser Eindruck täuscht: Tatsächlich handelt es sich um einen langen evolutionären Prozess, der jetzt zu einem qualitativen Umbruch führt.

## **28.1 Die Entwicklung von Interfaces zur Interaktion mit KI**

Diese Entwicklung lässt sich anschaulich an der Art und Weise ablesen, wie wir als Nutzer mit diesen Technologien interagieren. Vor etwa 25 Jahren, genauer gesagt vor eher 20 Jahren, wurde der Browser erfunden - ein technisches Hilfsmittel, mit dem wir aufbereitete HTML-Webseiten betrachten können. Der entscheidende Clou dieser am CERN entwickelten Technik bestand darin, dass die Seiten Verlinkungen zu anderen Seiten enthielten und so ein schnell wachsendes Netzwerk an Informationen bereitstellten. Vor rund 15 Jahren folgte dann die Erfindung des Smartphones, das heute aus unserem Alltag kaum noch wegzudenken ist und uns einen ähnlichen Zugriff auf Inf# Einführung in die Interaktion mit KI-Systemen

Heutzutage interagieren wir hauptsächlich über drei grundlegende Techniken mit künstlicher Intelligenz und den dadurch bereitgestellten Informationen. Die erste und wohl bekannteste Methode ist die Eingabe über ein Textfeld, das vor 25 Jahren mit dem HTTP-Protokoll definiert wurde. Dieses Feld, das oft fälschlicherweise als "Eingabefeld" bezeichnet wird, ermöglicht es Ihnen, mithilfe der Tastatur Links zu anderen Quellen einzugeben. Sie kennen diese Funktion vom Browser, wo Sie in der Adresszeile einen Link eintippen können.

## **28.2 Entwicklung der Eingabemöglichkeiten**

Ursprünglich diente das Adressfeld ausschließlich dazu, Verknüpfungen zu anderen Seiten und Adressen einzugeben. In den letzten 15 Jahren hat sich dieses Feld jedoch weiterentwickelt

und erlaubt nun die Eingabe von weiteren Anfragen – eine Funktion, die technisch gesehen gar nicht so anspruchsvoll ist, aber enorme Auswirkungen hat. Die berühmte Google-Suche ist ein perfektes Beispiel dafür: Anstatt selbst die Webadressen weiterer Quellen eingeben zu müssen, überlassen Sie diese Aufgabe nun einer Suchmaschine, die Ihre Anfrage beliebiger Art verarbeitet und Ihnen die entsprechenden Suchergebnisse zurückgibt.

### 28.3 Der Aufstieg von Chat-GPT

Mit der Einführung von Chat-GPT erleben wir einen massiven Umbruch in der Interaktion zwischen Mensch und Maschine. Chat-GPT, eine dialogorientierte Seite, die die Interaktion mit der KI ermöglicht, hat einen tieferen Grund für ihren Erfolg, den ich gleich noch erläutern werde. Zunächst mag es wie ein cleverer Marketing-Trick erscheinen, doch tatsächlich war es der erste höchst erfolgreiche Auftritt der KI-Modelle über eine solche Chat-Interaktion.

Wir befinden uns derzeit an einem Wendepunkt, an dem sich die Art und Weise, wie wir mit Maschinen interagieren, radikal verändert. Was vorher nur dazu diente, Inhalte von Providern bereitzustellen, wird sich jetzt zu einer Interaktion mit einem KI-Modell entwickeln. Sie werden nicht mehr mit einem Provider kommunizieren, sondern mit einem KI-Modell interagieren, das Ihre Informations- und Mitteilungsbedürfnisse steuert.

### 28.4 Weitere Interaktionsmöglichkeiten

Neben der Texteingabe gibt es noch weitere Möglichkeiten, mit KI-Systemen zu interagieren:

- Sprachbefehle und Spracheingaben, wie Sie sie von Siri kennen, ermöglichen es Ihnen, Befehle über das Mikrofon eines Computers einzugeben, die dann in entsprechende Befehlsstrukturen umgesetzt werden und eine Reaktion der Maschine auslösen.
- Datenbrillen und Headsets eröffnen neue Möglichkeiten der Interaktion. Obwohl ein erster Versuch von Google vor sechs Jahren aufgrund von Bedenken hinsichtlich der Privatsphäre scheiterte, planen nun alle größeren Firmen die Einführung solcher Geräte. Als Tourist könnten Sie beispielsweise vor einem Monument in Rom stehen und über die Brille Informationen zu dessen Erbauung und Geschichte abrufen. Oder Sie sitzen in der Oper und lassen sich eine Szene über den Ohrhörer erläutern.
- Gesten, sowohl taktile als auch sichtbare, können als Signale für die KI dienen. Bei vollständig gelähmten Personen gibt es sogar Implantate, die Hirnströme nutzen, um Signale nach außen zu senden.

Die Entwicklung in all diesen Bereichen schreitet rasant voran, und es bleibt spannend zu beobachten, in welche Richtung sie sich in den nächsten Jahren bewegen wird.

# 29 Die Architektur hinter den KI-Systemen

## 29.1 Generative KI

Auf den ersten Blick mag die Funktionsweise der Software und Programme im Hintergrund höchst kompliziert erscheinen – schließlich können derzeit nur die größten Konzerne mit Milliardenaufwand solche Modelle erstellen. Doch im Kern ist die Architektur gar nicht so kompliziert, wie wir gleich sehen werden.

Es geht hier um die sogenannte generative KI, oft auch abgekürzt als Gen-AI (nicht zu verwechseln mit dem Begriff “Gen”). Diese Systeme erzeugen etwas, das einen bedeutungsvollen sprachlichen Ausdruck darstellt – und genau das ist der revolutionäre Aspekt. Bisher bestanden die Techniken aus Zeichenfolgen, die lediglich eine bestimmte Regelhaftigkeit, eine Syntax, erfüllten, um als bedeutungsvolle Zeichenkette zu erscheinen. Ein Beispiel dafür wäre ein Satz, der im Deutschen durch einen Satzpunkt beendet und durch eine Großschreibung begonnen wird. Andere europäische Sprachen kennzeichnen die Syntax von Sätzen auf unterschiedliche Weise, aber darauf kommt es hier nicht an.

## 29.2 Von der Syntax zur Semantik

Bisher beschränkte sich der Umgang von Computern mit unserer sprachlichen Welt auf die Verarbeitung von Zeichenketten – auf die Syntax. Doch jetzt kommt etwas völlig Neues hinzu, und das ist die große Stunde der Philosophie: die Semantik.

- Die Syntax ist der sprachliche Ausdruck, die Zeichenketten, die Abfolge von Buchstaben, Wörtern und Sätzen, die linear verknüpft werden, um beispielsweise ein Buch zu bilden. All das sind sprachliche Ausdrücke oder, wie es der Philosoph Frege formulierte, der sinnlich wahrnehmbare Ausdruck sprachlicher, gedanklicher Inhalte.
- Die Semantik hingegen befasst sich mit der Bedeutung dieser Zeichen. Und genau damit haben wir es hier zum ersten Mal durch die Technik zu tun.

Im Gegensatz zu den vollmundigen Behauptungen der Konzerne, die schon von “Knowledge Graphen” à la Google sprachen, als von Bedeutung noch gar nicht die Rede war, sollte man diese Terminologie philosophisch hinterfragen. Dann wird schnell klar, dass das Kartenhaus

ziemlich schnell zusammenfällt. Es handelt sich nicht um "Knowledge Graphen", sondern um ganz einfache Graphen. Von Wissen ist da noch keine Spur.

Die philosophische Kritik an der Terminologie entlarvt, was hinter diesen Begrifflichkeiten eigentlich steckt. Und jeder, der bisher von der Bedeutung einer Aussage eines Computers gesprochen hat, weiß nicht, was es üblicherweise in der analytischen Philosophie bedeutet, von Bedeutung zu reden.<sup>#</sup> Die AI-Revolution: Sprache und Bedeutung

Meine Damen und Herren, heute möchte ich Ihnen eine faszinierende Entwicklung näherbringen, die unser Verständnis von Sprache und Bedeutung grundlegend verändern wird: die AI-Revolution. Im Kern geht es darum, dass künstliche Intelligenz nun in der Lage ist, sprachliche Ausdrücke mit ihrer Bedeutung zu verbinden. Diese Fähigkeit hat weitreichende Konsequenzen, die ich Ihnen heute andeuten möchte.

### **29.3 Von der Suche nach Zeichenketten zur Suche nach Inhalten**

Bisher waren Suchmaschinen wie Google darauf beschränkt, nach Zeichenketten zu suchen. Sie gaben einen Begriff ein und die Maschine suchte nach passenden Wörtern, Namen oder Adressen. Damit ließ sich schon viel erreichen, aber im Grunde war es nichts anderes als eine Suche nach Zeichenfolgen.

Doch nun eröffnet sich eine völlig neue Dimension: Die Suche nach den Inhalten und Aussagen, die mit sprachlichen Ausdrücken getroffen werden können. Lassen Sie mich das an einem einfachen Beispiel verdeutlichen:

Der Satz "Der Hund ist schwarz" ist zunächst einmal eine Zeichenkette. Doch diese Zeichenkette ist noch kein Inhalt. In der Philosophie unterscheiden wir streng zwischen dem Satz selbst und der Bedeutung, die er ausdrückt.

### **29.4 Sätze, Aussagen und Wahrheitswerte**

Sätze sind weder wahr noch falsch - eine Aussage, die manchen Informatikern vielleicht überraschend erscheinen mag. Sätze sind sprachliche Ausdrücke, die wohlgeformt sein können, aber keinen Wahrheitswert haben. Wahr oder falsch sind hingegen die mit Sätzen ausgedrückten Inhalte, die wir in der Philosophie als Aussagen oder Propositionen bezeichnen.

Solange wir uns nur auf der Ebene der Syntax bewegen, sind wir noch nicht einmal in der Welt des Wahren und Falschen angelangt. Und ohne Wahrheit oder Falschheit können wir auch nichts glauben oder für richtig halten. Überzeugungen entwickeln wir erst, wenn wir es mit etwas zu tun haben, das wahr oder falsch sein kann - eben mit Aussagen.

## 29.5 Die epistemische Dimension des Wissens

Aussagen sind die Träger von Wahrheitswerten. Und erst wenn wir von Aussagen mit Wahrheitswerten sprechen, kommen wir in die Sphäre der Rechtfertigung, der Kritik und der Widerlegung. Die epistemische Seite des Wissens - das Behaupten, Finden, Kritisieren und Widerlegen von Wissen - setzt voraus, dass wir es mit Aussagen und ihren Wahrheitswerten zu tun haben.

Eine Suchmaschine, die nur Zeichenketten findet, können Sie nicht kritisieren. Sie hat ihre Aufgabe erfüllt, auch wenn das Ergebnis vielleicht nicht Ihren Erwartungen entspricht. Kritik wäre hier fehl am Platz, ja geradezu ein Kategorienfehler.

## 29.6 Die Maschine lernt, Aussagen zu treffen

Wie aber gelingt es nun der Maschine, Aussagen zu treffen - eine Fähigkeit, die wir bisher nur dem menschlichen Geist, der Vernunft zugeschrieben haben? Die AI-Modelle werden derzeit mit höchstem Aufwand darauf trainiert,

1. bedeutungsähnliche Begriffe und Sätze zu unterscheiden,
2. den Strom der sprachlichen Zeichen in Wörter und Satzzeichen zu zerlegen (sogenannte Token),
3. und schließlich die Bedeutungsähnlichkeit von Sätzen zu erkennen.

Nehmen wir drei Beispiele:

- An eagle flies silently over the large tree.
- A swan flies noisily over the large tree.
- A mouse eats happily a piece of cheese.

Intuitiv erkennen wir sofort, dass die ersten beiden Sätze in ihrer Bedeutung ähnlich sind, auch wenn die Vögel verschieden sind und sich ihre Art zu fliegen unterscheidet. Wir würden sagen: Es fliegt ein Vogel auf eine bestimmte Weise über einen Baum.

Mit dieser Formulierung greifen wir automatisch auf die Bedeutung der Ausdrücke zu. Wir verallgemeinern soweit, dass wir von Vögeln sprechen, obwohl das Wort "Vogel" gar nicht vorkommt. Wir erkennen die semantische Ähnlichkeit der Sätze.

Der dritte Satz hingegen hat inhaltlich kaum etwas mit den ersten beiden gemeinsam. Allenfalls könnte man sagen, dass es auch hier um ein Tier geht. Aber sonst?

Genau diese Fähigkeit, Bedeutungsähnlichkeiten zu erkennen und Aussagen zu treffen, wird den AI-Modellen nun beigebracht. Und damit eröffnet sich eine völlig neue Dimension der Sprachverarbeitung, die weit über die bloße Suche nach Zeichenketten hinausgeht.<sup>#</sup> Einführung

Meine sehr verehrten Damen und Herren,

lassen Sie uns heute eine faszinierende Reise in die Welt der künstlichen Intelligenz und der Sprachverarbeitung unternehmen. Wir werden ergründen, wie es möglich ist, dass Maschinen die Bedeutung von Wörtern und Sätzen erfassen können - eine Fähigkeit, die lange Zeit als einzigartig menschlich galt.

## 29.7 Die Revolution der Sprachmodelle

Die erste Revolution auf diesem Gebiet ereignete sich, als man begann, die Sprachmodelle mit praktisch der Gesamtheit aller im Internet verfügbaren Texte zu trainieren. Wir sprechen hier von Trillionen von Worteinheiten, sogenannten Token, die als Grundlage dienten. Nicht irgendwelche speziellen Texte, sondern alles, was überhaupt im Netz zu finden ist, wurde herangezogen, um etwas zu definieren, das technisch gesehen als "Embedding" oder "Einbettung" bezeichnet wird.

### 29.7.1 Das Prinzip der Embeddings

Lassen Sie mich das Prinzip der Embeddings näher erläutern. Es handelt sich dabei um komprimierte Zahlenwerte, die Aufschluss darüber geben, in welchem Verwendungszusammenhang bestimmte Wörter mit anderen Wörtern stehen können. Im Grunde genommen werden gigantische Tabellen erstellt, die nichts anderes tun, als zu registrieren, welches Wort welchem anderen Wort folgt und in welchem Kontext von anderen Wörtern es auftritt.

Über mathematisch raffinierte Verfahren, auf die wir hier nicht näher eingehen müssen, lassen sich diese Tabellen so weit kombinieren und komprimieren, dass am Ende eine Tabelle mit 1500 Spalten ausreicht, um jedem einzelnen Satz eine Zuordnung zu geben, welche Rolle jedes Wort innerhalb dieses Satzes für die Bedeutung spielt. Das ist eine erstaunlich geringe Zahl, wenn man bedenkt, wie komplex Sprache ist.

### 29.7.2 Die Bedeutung eines Satzes

Oft wird vereinfachend gesagt, dass dieser Zahlenwert von 1536 Zahlen die Bedeutung eines Satzes ausdrückt. Das ist jedoch nicht ganz korrekt. Zunächst einmal drückt er nur die Kombinationshäufigkeit der Wörter untereinander aus - ein Schritt vor der eigentlichen Frage nach der Bedeutung. Aber es bringt uns der Antwort näher.

## **29.8 Die Herausforderung der Bedeutungsgleichheit**

Eine der ersten Herausforderungen, denen sich die KI-Forschung stellte, war die Frage, welche verschiedenen sprachlichen Ausdrücke die gleiche Bedeutung haben. Eine einfache Frage, die jedoch schwierig zu beantworten ist: Wie kann man maschinell erkennen, dass unterschiedliche Sätze, die von der Syntax her verschieden sind, dennoch das Gleiche ausdrücken?

### **29.8.1 Beispiele für Bedeutungsgleichheit**

Lassen Sie uns einige Beispiele betrachten:

- Die Aktiv-Passiv-Konvertierung: “Der Hund jagt die Katze” und “Die Katze wird vom Hund gejagt” bedeuten das Gleiche, obwohl sie sprachlich verschieden sind.
- Die Übersetzung: “Der Hund ist schwarz” hat die gleiche Bedeutung wie “The dog is black”. Obwohl jedes einzelne Wort vom Ausdruck her verschieden ist, drücken beide Sätze dasselbe aus.

## **29.9 Die Lösung durch künstliche Intelligenz**

In den letzten fünf Jahren hat die KI eine Lösung für diese Herausforderung gefunden. Wie Sie sich vorstellen können, war die Computerlinguistik schon seit mindestens 50 Jahren damit beschäftigt, dieses Problem zu lösen - allerdings mit nur mäßigem Erfolg. Doch jetzt ist es möglich, und das ist einer der Gründe, warum maschinelle Übersetzung heute zur Grundausrüstung von KI-Modellen gehört.

### **29.9.1 Komplexe Übersetzungen**

Moderne KI-Systeme sind in der Lage, komplexe Texte, sogar Fachtexte, adäquat in eine andere Sprache zu übersetzen. Und zwar nicht nur Wort für Wort, wie man es früher stümperhaft versucht hat, indem man jedes einzelne Wort in eine Übertragungstabelle einfügte und froh war, wenn die grammatischen Anforderungen halbwegs erfüllt wurden. Nein, heute kann ein Satz vollständig umgebaut oder sogar in Teilsätze zerlegt werden, um das Gleiche auszudrücken - genauso wie es ein guter menschlicher Übersetzer tun würde.

## **29.10 Das Training der KI-Modelle**

Embeddings sind die Grundvoraussetzung für diesen Erfolg. Sie sind ein Teil des riesigen Trainings, das die KI-Modelle durchlaufen. Für GPT 3.5 beispielsweise dauerte das Training etwa zwei Jahre und erforderte einen extremen Computeraufwand. Durch dieses Training anhand von Embeddings und vielen Textbeispielen lernen die KI-Modelle, die Frage nach Bedeutungsgleichheit erfolgreich zu beantworten.

### **29.10.1 Die Parameter der Modelle**

Das Training ist im Grunde ein Feintuning von Milliarden von Parametern - Stellschrauben, die so justiert werden müssen, dass die KI die Anforderungen an semantische Regeln richtig umsetzt. Das Trainingsziel ist dabei ganz einfach: Die Frage nach der Bedeutung zu lösen.

### **29.10.2 Trainingsdatensätze und Übersetzlitteratur**

Für das Training gibt es hervorragende Datensätze, an denen man den Erfolg messen kann. Einer der entscheidenden Datensätze sind die Klassiker der Übersetzlitteratur. Hier haben die besten Übersetzer der Welt eine literarische Quelle in einer Sprache vorgegeben und eine höchst anspruchsvolle Übersetzung als bedeutungsgleichen Ausdruck zugeordnet. Alles, was die großen Konzerne an Übersetzlitteratur bekommen konnten, haben sie für das Training verwendet.

Das ist eines der Geheimnisse, warum nun plötzlich auch Latein gut übersetzt wird. Die KI-Entwickler haben die Klassiker der Teutner-Serie genommen, die hervorragenden Übersetzungen der Philologen, und hatten damit eine präzise Übersetzungszuordnung zwischen modernen Sprachen und den Texten von Horaz oder Cicero.

## **29.11 Weitere Trainingsdaten**

Natürlich hat man auch die gesamte philosophische Literatur digitalisiert, denn hier finden sich wertvolle sprachphilosophische Reflexionen über die Inhalte. Was sind logische Schlussformen? Das kennen Sie alles aus den Logik-Lehrbüchern. Sie können aus den KI-Programmen herauskitzeln, dass sie diese Texte Satz für Satz trainiert haben. Nicht nur Philosophie-Studenten üben in Logik 1 die Logiktexte, auch alle KI-Modelle haben das intus, weil hier die Regeln der Semantik geübt werden. Ein Modus ponendo ponens gehört zum Repertoire des Schließens für KI-Modelle genauso wie für einen Philosophie-Studenten.

Allerdings geht das manchmal auch noch deutlich daneben...# Bedeutungsgleichheit und Embeddings in der KI

In der Welt der künstlichen Intelligenz spielen Wiederholungen eine ebenso entscheidende Rolle wie beim menschlichen Lernen. So wie ein einzelner Besuch einer Logikvorlesung nicht ausreicht, um die Materie vollständig zu beherrschen, benötigen auch Maschinen mehrfache Wiederholungen, um Inhalte zu verinnerlichen. Embeddings, numerische Repräsentationen von Wörtern oder Sätzen, dienen hierbei als Grundlage für das Training von KI-Modellen zur Bedeutungszuordnung. Doch Vorsicht: Embeddings allein reichen nicht aus, um die Bedeutung sprachlicher Ausdrücke vollständig zu erfassen.

## **29.12 Kontextabhängigkeit der Bedeutung**

Die Frage, ob Ausdrücke semantisch gleich sind, lässt sich in den meisten Fällen nicht pauschal beantworten. Der Kontext, in dem Sprache verwendet wird, spielt eine entscheidende Rolle bei der Beurteilung der Bedeutung sprachlicher Vorkommnisse. Embeddings reduzieren die komplexen Bedeutungsdimensionen auf einige Tausend mathematische Dimensionen - eine starke Vereinfachung, die jedoch den aktuellen Stand der Technik widerspiegelt.

## **29.13 Funktionsweise der KI bei inhaltlichen Fragen**

Stellen Sie sich vor, Sie fragen eine KI: "Fliegt da ein Schwan über den Baum?" Die KI übersetzt diese Eingabe zunächst in eine numerische Repräsentation, die sogenannten Embeddings. Dieser Satz erhält dann eine Zahl in 5.536 Dimensionen zugeordnet - eine erstaunlich kompakte Darstellung der dahinterstehenden Komplexität. Mit dieser Zahl durchsucht die KI eine Datenbank nach bedeutungsähnlichen Aussagen, unabhängig von der Sprache oder syntaktischen Transformationen wie Aktiv-Passiv-Konstruktionen. Die Zeichenabfolge (Strings) spielt keine Rolle mehr; es geht einzig um den Inhalt.

## **29.14 Erweiterung der Embeddings auf multimediale Inhalte**

Die Revolution der KI beschränkt sich nicht nur auf Texte. Embeddings lassen sich auch auf Bilder, Videos, Audio, 3D-Objekte und sogar Hologramme anwenden. KI-Programme können somit nicht nur Texte inhaltlich verstehen, sondern auch begleitende visuelle Elemente wie Diagramme oder Daten erschließen. Diese Erweiterung eröffnet völlig neue Möglichkeiten der Informationsverarbeitung.

# **30 Attention is all you need - die zweite Revolution**

Der Artikel “Attention is all you need”, erschienen auf dem Preprint-Server arXiv der Cornell University, markiert einen weiteren Meilenstein in der KI-Revolution. Die Autoren, darunter Jakob Uskoreits Sohn, haben sich ihr ganzes Leben mit Übersetzungen beschäftigt - ein Bereich, der den Boden für die KI-Revolution bereitet hat.

## **30.1 Transformation von Sequenzen**

Der Artikel befasst sich mit der Transformation von Sequenzen, also Satzabfolgen. Sprache wird hier als eine Abfolge von Satztokenwörtern verstanden. Die Aufgabe besteht darin, nicht nur die Bedeutung dieser Ausdrücke zu identifizieren, sondern auch vorherzusagen, welches Wort als nächstes folgen könnte. Diese Funktion kennen Sie vielleicht von Rechtschreibkorrekturprogrammen oder der Wortvervollständigung auf Smartphones.

## **30.2 Die Bedeutung der Sequenztransformation**

Die Fähigkeit, die nächste Zeichenfolge vorherzusagen, mag auf den ersten Blick technisch interessant, aber nicht besonders aufregend erscheinen. Doch genau hier liegt der Schlüssel zur zweiten Revolution. Denn es geht nicht nur um sprachliche Ausdrücke, sondern um die Dimension der Bedeutung.

# 31 Rettungsversuch und KI-Demonstration

Nachdem meine Folie sich unbeabsichtigt geschlossen hat, versuche ich einen Rettungsversuch mit Hilfe der KI. Nicht, um nach einer Lösung zu fragen, sondern um Ihnen live zu demonstrieren, was der Clou an der Vorhersage der nächsten Zeichenfolge ist.

Ich gebe den Satz “Der Hund ist schwarz.” ein. Was die KI als Fortsetzung vorschlägt, ist jedes Mal anders und oft überraschend. In der Evolution des User-Interfaces ist das, was früher der Go# Interaktion mit KI-Modellen

Heute möchte ich Ihnen von einem faszinierenden Experiment berichten, das ich kürzlich mit einem KI-Modell durchgeführt habe. Stellen Sie sich vor, Sie geben dem Programm eine einfache Aussage ein, wie beispielsweise “Der Hund ist schwarz.” Was erwarten Sie, dass das Modell darauf antwortet? Genau das habe ich ausprobiert und die Ergebnisse waren höchst aufschlussreich.

## 31.1 Die Herausforderung der Feststellung

Als ich die Aussage “Der Hund ist schwarz.” in das KI-Modell eingab, war ich gespannt, welche Reaktion es zeigen würde. Zu meiner Überraschung schien das Programm zunächst etwas perplex zu sein. Es wusste offenbar nicht so recht, was es mit dieser schlichten Feststellung anfangen sollte.

Ich wiederholte die Eingabe, doch das Modell reagierte erneut mit einer Entschuldigung und gab dann eine ausführliche Beschreibung eines schwarzen Labradors aus. Es assoziierte alles, was man mit schwarzen Hunden in Verbindung bringen könnte und generierte einen regelrechten pseudoliterarischen Erguss.

Als ich die Aussage ein weiteres Mal wiederholte, erkannte das Programm immerhin, dass seine vorherige Antwort wohl nicht ganz das Richtige war. Es entschuldigte sich erneut und wiederholte die ursprüngliche Aussage in ihrer prägnanten Form.

## **31.2 Von der Frage zur Anweisung**

Dieses Experiment verdeutlicht einen wichtigen Wandel in der Nutzung von KI-Modellen. Anfangs dienten sie hauptsächlich dazu, Fragen zu beantworten - ähnlich wie bei einer Google-Suche. Die Pragmatik des Dialogführers war klar: Frage und Antwort.

Doch mittlerweile hat sich der Fokus verschoben. Statt Fragen zu stellen, geben wir den Modellen immer häufiger Anweisungen oder Instruktionen. Deshalb wurden viele Modelle neu trainiert und tragen nun Namen, die auf ihre Fähigkeit zur Ausführung von Anweisungen hindeuten.

## **31.3 Die Bedeutung der Aufmerksamkeit**

Der Attention-Mechanismus spielt bei der Generierung plausibler Textfolgen eine entscheidende Rolle. Je nachdem, ob wir eine Frage stellen, eine Instruktion geben oder etwas sagen, das offensichtlich eine bestimmte Reaktion erfordert, passt sich die Ausgabe des Modells an.

Doch was passiert, wenn wir dem Programm eine Aussage präsentieren, auf die es keine sinnvolle Antwort geben kann? Hier zeigt sich eine interessante Eigenschaft der meisten KI-Modelle: Sie sind so programmiert, dass sie immer etwas ausgeben müssen. Schweigsame Modelle, die bei einer Unsinnfrage einfach nichts sagen, gibt es nicht.

## **31.4 Die Macht der Kontextualisierung**

Der revolutionäre Aspekt der gegenwärtigen KI-Modelle liegt in ihrer Fähigkeit, sprachliche Ausdrücke zu kontextualisieren. Nehmen wir das Beispiel der Übersetzung. Wenn ich dem Programm die Anweisung gebe: “Übersetze den Text ‘Der Hund ist schwarz’”, dann versteht es die Bedeutung und gibt korrekt “The dog is black” aus.

Dieser Komplex aus Anweisung und sprachlichem Ausdruck wird vom Modell richtig verstanden und die entsprechende Ausgabe generiert. Intern reformuliert das Programm die Eingabe in eine explizite Wiedergabe des Inhalts, um sicherzustellen, dass alles eindeutig ist. # Erweiterung des Eingabekontexts zur Steuerung der Ausgabe

In den gegenwärtigen KI-Modellen wird der Kontext explizit gemacht und mit in den Eingabekontext geschrieben. Auf diese Weise wird der generierte Text gesteuert. Wenn ich beispielsweise ohne weiteren Kontext den Satz “Der Hund ist schwarz” eingebe, fängt das Programm von sich aus an, weitere Informationen zu generieren. Durch Zusatzinformationen im Kontext lässt sich die Ausgabe jedoch stark beeinflussen.

## **31.5 Das Problem der Halluzination**

Die Halluzination entsteht dadurch, dass es keinerlei Beschränkungen auf den Inhalt oder sachliche Prüfungen gibt. Die aktuellen Modelle beherrschen lediglich die Übersetzung von sprachlichem Ausdruck in ihre Bedeutung - sie haben Sprachkompetenz, aber keinerlei Sachkompetenz. Es existieren keine Mechanismen, die prüfen, ob ein generierter Satz tatsächlich sachlich korrekt ist.

Obwohl durch die KI-Programme die Dimension der Wahrheit, Rechtfertigung und Kritik eröffnet wird, lösen sie diese Fragen noch nicht ein. Sachliche Korrektheit wird nicht geprüft, Evidenzen nicht angeführt und Kritik nicht geübt. Diese Aspekte sind schlichtweg nicht Teil der Programme.

## **31.6 Konsequenzen für die Verwendung von KI-generierten Texten**

Hausarbeiten sollten niemals mit Chat-GPT geschrieben werden, da die Wahrscheinlichkeit für falsche Informationen extrem hoch ist. Die Programme sind perfekt darin, Ausgaben sinnvoll erscheinen zu lassen, aber nicht in der Lage, deren Wahrheitsgehalt zu überprüfen.

Ein Beispiel hierfür ist meine Erfahrung mit einer Abfrage zu Leonhard Eulers Publikationsverhalten im Jahr 1756. Statt zuzugeben, keine Informationen zu haben, generierte das Programm eine perfekt aussehende, aber völlig erfundene Literaturangabe. Selbst als Experte konnte ich die Fälschung zunächst nicht erkennen - so überzeugend war die Formatierung bis hin zu passenden bibliografischen Details. Ohne Fachwissen wäre der Fake nicht aufgefallen.

# **32 Erweiterung der KI-Modelle um Wissen und Validierung**

## **32.1 Notwendige Ergänzungen für sachliche Korrektheit**

Um zu garantieren, dass generierte Informationen richtig sind, müssen den KI-Modellen zusätzliche Elemente hinzugefügt werden. Sie benötigen Zugriff auf das Wissen der Welt, das ihnen momentan fehlt.

Als Wissenschaftler würde man zur Prüfung einer Aussage wie folgt vorgehen:

- Konsultation glaubwürdiger Referenzen
- Recherche in Primärquellen (z.B. Eulers Opera Omnia)
- Aufsuchen einer Bibliothek zur Verifizierung der Publikation

Dieses Vorgehen müsste in zukünftigen KI-Systemen abgebildet werden, um sachliche Korrektheit herzustellen.

## **32.2 Verhältnis von Sprache und Sachlichkeit**

Sprache und Sachlichkeit sind vergleichbar komplex. Der sprachliche Ausdruck sollte idealerweise dem sachlichen Inhalt in der Welt entsprechen - eine Korrespondenztheorie der Wahrheit. Stimmen beide überein, ist die Aussage wahr, ansonsten falsch.

Diese Korrespondenz muss den KI-Modellen methodisch beigebracht werden, um über reine Sprachkompetenz hinauszugehen. Aktuell fehlt ihnen der Zugriff auf die Realität, auf die sich die Sprache beziehen sollte.

# **33 Auswirkungen der KI-Entwicklung auf Sprache und Bedeutung**

## **33.1 Zirkularität der Bedeutung in KI-trainierten Texten**

Wenn immer mehr Texte von KI-Systemen generiert werden, die auf jahrhundertealten Daten trainiert wurden, entsteht die Gefahr einer Zirkularität der Bedeutung. Neue Bedeutungsebenen von Begriffen könnten nicht mehr hinzukommen, die Sprache käme zum Stillstand - zumindest bei Systemen, die nur reproduzieren, was sie in den Vorlagen finden.

## **33.2 Übersetzungsfähigkeiten aktueller Programme**

Die aktuellen Programme arbeiten bereits auf der Bedeutungsebene und sind in der Lage, beliebige Sätze zu übersetzen, auch wenn die übersetzte Formulierung nirgends in der Literatur vorhanden ist.

Selbst anspruchsvolle Texte wie Goethes Faust oder Werke von Thomas Mann, die nicht in jede Sprache übersetzt wurden, können von den Programmen übertragen werden. Ob die Übersetzung angemessen ist oder Fehler enthält, lässt sich diskutieren - aber die Programme werden einen Übersetzungsvorschlag liefern. # Einleitung

Einen schönen guten Morgen, meine Damen und Herren. Heute möchte ich Ihnen etwas über die faszinierenden Entwicklungen im Bereich der Künstlichen Intelligenz und insbesondere der Sprachmodelle erzählen. Lassen Sie uns gemeinsam ergründen, wie diese Systeme funktionieren und welche Herausforderungen und Möglichkeiten sich daraus ergeben.

## 34 Die Bedeutung der Sprachverwendung

Zunächst einmal stellt sich die Frage, wie die Bedeutung in der Sprache eigentlich entsteht. In der Philosophie gibt es dazu verschiedene Ansätze, aber eine zentrale Erkenntnis ist, dass die Bedeutung eng mit der tatsächlichen Verwendung der Sprache verknüpft ist. Die Sprachmodelle der KI versuchen genau das zu erfassen - sie analysieren riesige Textkorpora und lernen daraus, in welchen Kontexten bestimmte Ausdrücke typischerweise vorkommen.

Aber bedeutet das nun, dass die Entwicklung der Sprachmodelle davon abhängt, dass immer mehr Texte produziert werden? Nein, so einfach ist es nicht. Die Trainingsdaten sind in der Regel nur eine repräsentative Teilmenge aller verfügbaren Texte. Und natürlich verändert sich Sprache auch im Laufe der Zeit. Die historische Entwicklung von Bedeutungsverschiebungen ist eine große Herausforderung für die Forschung. Hier gibt es noch viel zu tun.

## 35 Die Gefahren fehlerhafter Kontexte

Ein interessantes Phänomen, das wir beobachten konnten, sind Sprachmarotten, die in den Modellen entstehen können. In einem Fall wurden die Modelle offenbar mit Texten trainiert, in denen es nicht um tatsächliche Kausalzusammenhänge ging, sondern darum, was Personen glauben, was die Ursache von etwas ist. Das führte dazu, dass die Modelle Unsinn produzierten, wenn es um kausales Schließen ging.

Dieses Beispiel zeigt, wie wichtig die sorgfältige Auswahl und Aufbereitung der Trainingsdaten ist. Fehlerhafte oder irreführende Kontexte können sich hartnäckig in den Modellen festsetzen. Manchmal hilft es schon, in den Eingabetexten explizit klarzustellen, welche Art von Antwort man erwartet. Aber in manchen Fällen muss man vielleicht auch einsehen, dass das Modell für bestimmte Aufgaben einfach nicht geeignet ist.

## **36 Erfolge und Anwendungen**

Trotz dieser Herausforderungen gibt es aber auch beeindruckende Erfolge zu vermelden. Übersetzungen waren nicht nur ein kultureller Gewinn, sondern auch ein wichtiger Motor für das Training von Bedeutungsgleichheit. Die Modelle sind inzwischen in der Lage, hochwertige Zusammenfassungen von langen Texten zu erstellen. Mit einer Kontextlänge von 200.000 Wörtern kann man ganze Bücher eingeben und sich die Kapitel in kompakter Form zusammenfassen lassen.

## 37 Die Bedeutung des Chats

Ein faszinierender Aspekt, der oft übersehen wird, ist die Rolle des Chats in der Interaktion mit Sprachmodellen wie ChatGPT. Die Entwickler wissen natürlich, dass man Begriffe nicht einfach durch notwendige und hinreichende Definitionen eingeben kann. Stattdessen nutzen sie Wittgensteinsche Gebrauchsdefinitionen. Und genau hier kommt der Chat ins Spiel.

In einem Chat findet oft eine Art semantische Korrektur statt. Wir fragen "Was meinst du damit?" oder "Meinst du dies oder jenes?". Dadurch klären wir die Bedeutung dessen, was der andere gesagt hat. Und genau das passiert auch in der Interaktion mit ChatGPT. Wenn wir eine Frage stellen und das Modell antwortet, dann können wir durch Rückfragen und Klarstellungen den Kontext präzisieren.

Das bedeutet, dass wir als Nutzer aktiv zur Intelligenz des Systems beitragen. Indem wir chatten, helfen wir dem Modell, die Bedeutung zu erschließen und bessere Antworten zu geben. Das ist ein genialer Ansatz, der bewusst so entworfen wurde und bis heute genutzt wird.

# **38 Philosophie der AI**

ai\_Vorl3

# **39 Begrüßung und Einführung in die Vorlesung “Philosophie der AI”**

Herzlich willkommen zur ersten Vorlesung “Philosophie der AI”! Ursprünglich trug diese Veranstaltung den Titel “Philosophie der künstlichen Intelligenz”, doch angesichts der aktuellen Diskussionen habe ich mich entschieden, den Begriff auf “AI” zu verkürzen. In diesem Semester möchte ich Ihnen einen umfassenden Überblick über die philosophischen Beiträge und Fundamente der modernen Artificial Intelligence geben und Sie durch die Grundlagen führen.

## **39.1 Die Rolle der Philosophie in der AI**

Entgegen der Erwartungen vieler geht es in dieser Vorlesung nicht primär darum, eine Bewertung oder Reflexion über die Folgen und Konsequenzen der künstlichen Intelligenz vorzunehmen. Obwohl wir diese Themen en passant ebenfalls behandeln werden, liegt der Kern der Vorlesung in der Grundthese, dass die eigentliche Innovation und der technologische Kern hinter dem Funktionieren der KI nicht nur in der Informatik, Technologie oder der fortschreitenden Entwicklung der Gerätschaften und Chips liegt, sondern in der Philosophie selbst. Ich vertrete die Ansicht, dass die künstliche Intelligenz heute eine Renaissance der analytischen Philosophie zur Folge hat, die die eigentliche inhaltliche und systematische Basis dessen bildet, was wir heute unter KI verstehen. Es handelt sich hierbei um eine anspruchsvolle Position, die die Philosophie nicht nur als Kommentator der technologischen und gesellschaftlichen Entwicklungen betrachtet, sondern als essenziellen Teil dieser Bewegung und Entwicklung.

# **40 Die KI-Revolution und ihre Auswirkungen**

## **40.1 Eine technologisch-gesellschaftliche Revolution**

Wir befinden uns derzeit nicht nur inmitten einer technologisch-gesellschaftlichen, politischen und sonstigen Revolution, die in ihrer Tragweite mit der Einführung der Elektrizität vor 150 Jahren oder des Webs vor etwa 25 Jahren vergleichbar ist. Vielmehr stehen wir gerade am Anfang einer Phase der technologischen Revolution durch die Einführung der künstlichen Intelligenz, deren weitreichende Entwicklungen wir nur erahnen können. Ein Indiz dafür ist die Tatsache, dass technologische Veränderungen, Möglichkeiten und Nutzungsformen mittlerweile auf täglicher Basis geschehen.

## **40.2 Herausforderungen in der Vorlesungsvorbereitung**

Während der Vorbereitung dieser Vorlesung ist mir aufgefallen, dass man nicht davon ausgehen kann, mit denselben Utensilien, Tools und Hilfsmitteln zu beginnen und am Ende der Vorlesung weiterzuarbeiten. Die Möglichkeiten und technologischen Anforderungen ändern sich so rasant, dass sie sich sogar während des Verlaufs dieser Vorlesung weiterentwickeln werden. Mein Ziel ist es, Ihnen die Gelegenheit zu bieten, einige dieser Tools während der Vorlesung, in der Nachbereitung oder Vorbereitung selbst auszuprobieren.

# **41 Vorkenntnisse und Erwartungen an die Studierenden**

## **41.1 Vertrautheit mit ChatGPT**

Ich bin mir sicher, dass ein Großteil von Ihnen bereits mit Tools wie ChatGPT vertraut ist oder sich eingehender damit beschäftigt hat. Der Begriff ChatGPT dürfte Ihnen kein Fremdwort sein und Sie wissen, wie man damit umgeht. In einer vorbereitenden Vorlesung im letzten Semester an der LMU München war ich erstaunt, als ich feststellte, dass bereits vor einem halben Jahr praktisch die gesamte Studierendenschaft mit ChatGPT vertraut war und es nutzte. Daher werden Sie in dieser Vorlesung keine Einführung in ChatGPT erwarten können, sondern ich setze diese Kenntnisse voraus. Stattdessen werde ich versuchen, tiefer in die philosophischen Aspekte der künstlichen Intelligenz einzutauchen.

## **41.2 Begriff der AI oder KI**

Bevor wir uns den Inhalten zuwenden, möchte ich Sie fragen: Was verstehen Sie unter AI? Lassen Sie uns kurz darüber nachdenken, bevor wir fortfahren.

## **42 Organisatorisches und Tools**

### **42.1 Vorlesungszeiten und -ort**

Die Vorlesung beginnt, obwohl in Agnes als 00 angekündigt, tatsächlich um CT. Diese Änderung ist nicht auf meinen Mist gewachsen, sondern geht auf die HU-Verwaltung zurück. Üblicherweise beginnt die Vorlesung um Viertel nach 10 und endet um Viertel vor 12, damit Sie ausreichend Zeit haben, zwischen den Veranstaltungen oder Universitäten zu wechseln.

### **42.2 Vorlesungswebseite und Materialien**

- In spätestens zwei Wochen werde ich eine eigene Webseite zur Vorlesung aufschalten. Ich verzichte auf die Nutzung von Moodle, da es für Lehrende ein Folterinstrument und eine Zeitverschwendungen darstellt.
- In der nächsten Woche werde ich Ihnen den Link zur Webseite hier zur Verfügung stellen.
- Auf der Webseite finden Sie für jede Vorlesung eine mit KI verfasste Zusammenfassung zum Herunterladen sowie weitere Materialien und gegebenenfalls Verlinkungen.

### **42.3 Zugriff auf Chat-GPT**

- OpenAI hat mir vorvergangene Woche mitgeteilt, dass sie die Zugriffe zu Chat-GPT freigeben. Eine Anmeldung sollte nicht mehr erforderlich sein.
- Ich konnte dies selbst nicht ausprobieren, da ich einen POE-Account besitze und eine Rückstufung zu kompliziert war.
- Ich würde mich über eine Rückmeldung von Ihnen freuen, ob der Zugriff bei Ihnen funktioniert.

### **42.4 Eigene KI-Webseite und virtueller Wittgenstein**

- In der zweiten Hälfte der Vorlesung werde ich möglicherweise eine eigene Webseite mit neuen KI-Möglichkeiten, die ich mit meinem eigenen Lab entwickle, freischalten und Ihnen zugänglich machen.

- Unter anderem planen wir, mit einem revitalisierten Wittgenstein mittels KI zu philosophieren.
- Ich hoffe, dass wir Ihnen einen virtuellen KI-Wittgenstein präsentieren können, mit dem Sie nicht nur Texte austauschen, sondern auch philosophische Diskussionen führen können.
- Weitere Details dazu werden wir in der zweiten Hälfte der Vorlesung erfahren.

## 42.5 Moodle und Teilnehmerliste

- Das Passwort für die Moodle-Vorlesung lautet “1234”. Dafür benötigen Sie keine künstliche Intelligenz, sondern lediglich ein gutes Gedächtnis.
- Bitte tragen Sie sich in die Teilnehmerliste ein.
- Im Falle von terminlichen Schwierigkeiten oder unvorhergesehenen Ereignissen wie Streiks, die mich am rechtzeitigen Erscheinen hindern, möchte ich Sie gerne per E-Mail erreichen können. Dies ist nur möglich, wenn Sie sich auf Moodle mit Ihrer E-Mail-Adresse registrieren.

## 42.6 Zulassung und Modulabschlussnoten

- ÜWP-Studierende, die durch unser automatisches Nicht-Intelligenz-System Agnes abgelehnt wurden, sollten sich nicht beunruhigen. Solange ausreichend Plätze vorhanden sind, lasse ich grundsätzlich alle Interessierten zu.
- Sollten Sie jedoch Modulabschlussnoten und Bescheinigungen benötigen, informieren Sie mich bitte im Vorfeld darüber, da es hierbei administrative Begrenzungen gibt, die Ihre Anwesenheit erfordern.

## **43 Nachfragen und individuelle Anliegen**

Ich stehe Ihnen am Ende der Vorlesung gerne für Nachfragen, Scheinanforderungen und ähnliche Anliegen zur Verfügung, solange ich nicht aus dem Vorlesungssaal geworfen werde. Um die wertvolle Vorlesungszeit optimal zu nutzen, möchte ich diese Themen nicht während der Veranstaltung behandeln. Individuelle Fragen beantworte ich auch gerne per E-Mail, wenn Sie mir schreiben. Liebe Studierende, zu Beginn der heutigen Vorlesung möchte ich ein paar organisatorische Dinge ansprechen. Wie Sie sich zu bestimmten Zeiten in Register eintragen und welche Erfordernisse für eine Prüfung als ÜWP-Studierende nötig sind, hängt von den verschiedenen Fakultäten ab. Bitte erkundigen Sie sich individuell darüber. Prinzipiell ist die Vorlesung natürlich offen und ich lasse alles zu, soweit es mir möglich ist. Wir müssen nur auf die administrativen Regelungen achten.

## 44 Was ist AI?

Künstliche Intelligenz, oder kurz AI, ist ein Begriff für eine technische Möglichkeit, die Mitte der 50er Jahre von einigen Kollegen aus Pittsburgh erdacht wurde. Ihr Ziel war es, maschinelle Computertechnologien zu entwickeln, die den menschlichen kognitiven Fähigkeiten nicht nur ebenbürtig sind, sondern sie sogar übertreffen. Man versprach damals vollmundig, dass dieses ehrgeizige Ziel in nur drei bis vier Jahren erreicht sein würde. Die Menschheit könnte dann endlich ihre Freizeit in vollen Zügen genießen, nur noch wenige Stunden pro Woche arbeiten, während der Rest von der KI erledigt würde.

Doch wie wir alle wissen, hat sich von dieser Vision bisher nichts eingelöst. Die Vorstellung war, dass KI als Meisterdisziplin des menschlichen Denkens schnell alle Bereiche überflügeln würde. Als Paradebeispiel galt damals das Schachspiel. Doch erst Anfang der 2000er Jahre gelang es einem Computerprogramm, den Schachweltmeister Garri Kasparov in einem ernsthaften Spiel zu besiegen - immerhin 50 Jahre später als ursprünglich prophezeit.

Das andere große Ziel, Computer zu entwickeln, die selbstständig wissenschaftlich kreativ denken können, ist bis heute nicht wirklich erreicht. Trotz aller anderslautenden, manchmal sensationsheischenden Meldungen bin ich jedoch sicher, dass diese Stufe in den nächsten Jahren erreicht werden wird. Dass also wissenschaftliche, kreative, kognitive und intellektuelle Aktivitäten von Maschinen alleine, ohne Assistenz von Forschern gemeistert werden. Das ist sozusagen noch die Krönung der Herausforderung von KI, von Artificial Intelligence.

## 45 KI als Verkaufsargument

Was Ihnen derzeit tagtäglich in der Öffentlichkeit als KI präsentiert wird, hat mit den eigentlichen Visionen und Zielen oft wenig zu tun. Nehmen wir als Beispiel eine Anzeige der Firma Samsung für ihre “Bespoke AI 11-Kilogramm-Washing-Maschine Serie 8 mit AI-Eco-Bubble und Quick-Drive”. Technisch gesehen handelt es sich schlicht um eine Waschmaschine, aber das Label “AI” soll den Verkauf ankurbeln.

Was ist daran nun wirklich AI? Nicht viel, es ist mehr ein Verkaufsargument als alles andere. Alles, was halbwegs gesteuert ist, wird heutzutage als AI vermarktet. Wenn ich hier “Licht aus” sage und es dunkel würde, würden Sie vielleicht denken “Oh, wir haben AI an der HU”. Dabei ist es letztlich nur eine etwas anspruchsvollere Steuerungstechnik, mehr nicht. Das Wort AI ist hier fehl am Platz, auch wenn es gerade en vogue ist.

## 46 Der Durchbruch der KI-Visionen

Sind wir also jetzt in einer Zeit angekommen, in der sich die ursprünglichen KI-Visionen doch noch erfüllen könnten? Meine Antwort lautet: Ja. Und ich möchte Ihnen heute einen systematischen Grund dafür nennen, der für mich entscheidend ist und den ich Ihnen so vermitteln möchte, dass er nachvollziehbar wird. Nebenbei bemerkt: Wenn Sie Fragen oder Zwischenfragen haben, melden Sie sich einfach. Dann gestalten wir die Vorlesung etwas lebendiger und interaktiver.

Der Aspekt, auf den ich hinaus möchte und den ich für den Meilenstein halte, ist, dass die KI-Visionen gerade dabei sind Wirklichkeit zu werden. Die KI-Propaganda hingegen, die sollten wir schnell beiseite legen. Das ist in erster Linie ein Verkaufsargument, das nicht den Kern der technologischen Innovation ausmacht. Und genau das soll heute unser Thema sein.

# **47 Die Attraktivität von KI**

Wo liegt denn potenziell die Attraktivität der KI, wie immer wir uns ihr auch nähern? Ist es eine bessere Internetsuchmaschine, die derzeit vielleicht eine der Triebfedern ist? Um das zu verstehen, müssen wir uns die Entwicklung des Internets vor Augen führen.

Gemessen an der Technologiegeschichte ist das Internet noch gar nicht so alt, etwas mehr als 20 Jahre. Wer die Anfänge noch miterlebt hat, erinnert sich an die ersten Browser, die damals oft mit Duschanlagen verwechselt wurden. Vor 20 Jahren wussten die wenigsten, was ein Internetbrowser eigentlich ist. Mittlerweile können wir uns ein Leben ohne Internet kaum noch vorstellen, weder technisch noch gesellschaftlich.

## **47.1 Die ursprüngliche Idee des Internets**

Im Kern war die Konstruktion des Internets, die am CERN entwickelt wurde, folgende: Irgendwo stellen wissenschaftliche Einrichtungen webzugängliche Seiten als Informationsquellen bereit. Als Wissenschaftler oder technologische Provider verantworten sie die Inhalte, pflegen sie und sorgen für dauerhafte Zugänglichkeit. Die Browser sind lediglich das lesende Frontend für diejenigen, die auf die Inhalte zugreifen wollen.

Damals war das Internet also eine Art anspruchsvolles Faxgerät als Empfänger der Inhalte. Der Clou lag darin, dass man ganz einfach andere Inhalte per Verlinkung einbinden konnte. So entwickelte sich ein Schneeballsystem, das ein globales Netz von miteinander verknüpften Inhalten erzeugte. Das war die Webrevolution vor 20 Jahren.

## **47.2 Die Ablösung der Webwelt durch KI**

Was wir jetzt erleben, ist eine Ablösung dieser Webwelt durch KI. In den nächsten Monaten werden Sie zunehmend feststellen, dass nicht mehr die Provider die Netzinhalte erstellen, auf Webservern bereitstellen und per Browser zugänglich machen. Diese Grundarchitektur wird abgelöst. Nicht mehr der Browser verantwortet, pflegt und stellt die Inhalte bereit. Das ist eine revolutionäre Änderung der Architektur der Informationsflüsse, aber auch der damit verbundenen Probleme. Einen Teil davon werden wir noch kennenlernen oder haben Sie schon erfahren.

Das Web funktionierte bisher deshalb, weil die Inhalte von den jeweiligen Personen, Institutionen oder Wissenschaftlern, die sie bereitstellten, auch autorisiert wurden. Für die Korrektheit und Richtigkeit bürgten die Glaubwürdigkeit und Gewissenhaftigkeit der Provider. Das ändert sich jetzt. Und wir alle wissen um die Gefahren, aber auch Potenziale, die damit einhergehen.

- Auf der einen Seite sind es nun große Internetfirmen, die die Inhalte über KI-Maschinen, sogenannte Bots, bereitstellen.
- Auf der anderen Seite können es auch böswillige Gestalten, Institutionen oder Staaten sein, die Inhalte generieren, ins Netz einspeisen, ohne als autorisierende Internetprovider in Erscheinung zu treten.

Derzeit wird das unter dem Stichwort "Internetinhalte der Social Media" diskutiert. Doch das ist nur die Oberfläche. Der Kern des Wandels und des Problems liegt darin, dass die Grundarchitektur des Internets mit den verantwortlichen Providern abgelöst wird durch - ich will nicht sagen unverantwortliche Bots - aber zumindest durch nicht mehr verantwortliche Internetinhaltsprovider. Und das hängt eben mit der KI-Revolution und dem Wandel der Informationsflüsse im Internet zusammen.<sup>#</sup> Die Veränderung der Informationssuche im Zeitalter der Künstlichen Intelligenz

Meine sehr verehrten Damen und Herren, lassen Sie uns heute gemeinsam einen Blick in die Zukunft der Informationssuche werfen. Bislang war es für uns alle selbstverständlich, dass wir bei der Suche nach Informationen auf die Dienste von Suchmaschinen wie Google zurückgreifen konnten. Wir vertrauten darauf, dass die von diesen autoritativen Anbietern bereitgestellten Inhalte glaubwürdig und sorgsam kuratiert waren. Doch in der nächsten Phase der digitalen Revolution wird sich dies grundlegend ändern.

### **47.3 Die Umgestaltung der Architektur des Internets**

Die Architektur des Internets befindet sich in einem extrem dynamischen Wandlungsprozess, dessen Ausgang noch niemand vorhersehen kann. Eines ist jedoch sicher: Es werden enorme Anstrengungen unternommen und gewaltige finanzielle Mittel investiert, um diese Transformation voranzutreiben. Jeder Staat, jede Region und auch Europa sollte ein vitales Interesse daran haben, die Kontrolle über diese Entwicklung nicht zu verlieren.

### **47.4 Neue Möglichkeiten durch Künstliche Intelligenz**

Doch lassen Sie uns zunächst einen Blick auf die vielversprechenden Möglichkeiten werfen, die uns die Künstliche Intelligenz eröffnet. Vielleicht erscheinen Ihnen einige dieser Anwendungen auf den ersten Blick trivial, doch ich versichere Ihnen, sie haben das Potenzial, unseren Alltag und unsere Arbeit grundlegend zu verändern.

#### **47.4.1 Hochwertige Übersetzungen**

Nehmen wir zum Beispiel das Thema Übersetzungen. Seit Jahrzehnten wurden enorme Ressourcen in die Entwicklung von linguistischen Modellen zur automatischen Übersetzung von Sprachen investiert. Doch lange Zeit waren die Ergebnisse bestenfalls als Partygags zu gebrauchen und keinesfalls für den ernsthaften Einsatz geeignet. In den letzten Jahren hat sich dies jedoch grundlegend geändert. Mittlerweile sind die automatischen Übersetzungen von so hoher Qualität, dass sie sogar für akademische Zwecke genutzt werden können.

Lassen Sie mich Ihnen ein Beispiel aus meinem eigenen Fachgebiet, der Wissenschaftsgeschichte, geben. Viele der historischen Quellen, mit denen wir arbeiten, sind in Latein verfasst. Vor 100 Jahren mussten Doktoranden ihre Dissertationen an unserer Fakultät noch auf Latein einreichen. Heute würden die meisten von Ihnen wohl Schwierigkeiten haben, einen lateinischen Quelltext sinnvoll zu interpretieren. Doch dank der Fortschritte in der Künstlichen Intelligenz gibt es Hoffnung. Vielleicht führen wir ja in unserer Fakultät bald wieder die Pflicht ein, Doktorarbeiten auf Latein zu verfassen - mit KI als Hilfsmittel könnte dies durchaus ein Alleinstellungsmerkmal unserer Universität werden.

#### **47.4.2 Simultanübersetzung und Lektoratsassistentz**

Die Möglichkeiten gehen jedoch noch weiter. In naher Zukunft werden wir in der Lage sein, hervorragende Simultanübersetzungen anzubieten. Ausländische Studierende, die keine europäische Sprache beherrschen, könnten meine Vorlesung mit einem Ohrhörer verfolgen und eine simultane Übersetzung erhalten.

Auch im Bereich des Lektorats gibt es spannende Entwicklungen. Programme wie Grammarly oder DeepL Write bieten bereits heute Textverbesserungsvorschläge, die durchaus mit der Qualität professioneller Lektoratsassistenzen mithalten können. Selbst große wissenschaftliche Verlage wie Nature stellen ihren Autoren mittlerweile Tools zur Verfügung, um ihre englischen Texte in lesbare Form zu bringen. Ob und wie dies gewünscht ist, wird derzeit heiß diskutiert. Doch ich bin davon überzeugt, dass in Zukunft das KI-gestützte Lektorat für wissenschaftliche Publikationen zum Standard werden wird.

#### **47.4.3 Automatisierte Forschungsberichte**

Vor der Tür stehen bereits Modelle, die in der Lage sind, eigenständig Texte wie Forschungsberichte zu verfassen. In experimentellen Wissenschaften wie der klinischen Forschung wird bereits daran gearbeitet, Ergebnisse und Erkenntnisse automatisch in Berichte zu überführen, die qualitativ den gängigen Publikationen entsprechen. Dies wirft natürlich Fragen auf:

- Wer ist der Autor eines solchen Berichts?
- Akzeptieren wissenschaftliche Journals Texte, die von einer KI erstellt wurden?

- Wie gehen wir mit Verantwortlichkeit, Seriosität und Zurechenbarkeit um?

Diese Probleme müssen gelöst werden, wenn wir diese Entwicklung weiter vorantreiben wollen.

## 47.5 Das Labor für gebildete KI

In meinem eigenen Labor, Lettre AI, setzen wir die Techniken um, die ich Ihnen in dieser Vorlesung vermitteln möchte. Unser Ziel ist es, eine KI bereitzustellen, die über die Fähigkeiten des Lesens, Übersetzens und Formulierens hinaus auch epistemische Qualifikationen mitbringt - also wissenbezogene Fähigkeiten, die wir gleich noch näher kennenlernen werden.

Lassen Sie mich Ihnen ein Beispiel für die bereits existierende Leistungsfähigkeit von KI geben. Ich zeige Ihnen hier einen Ausschnitt aus einem Werk, das zu Beginn des 17. Jahrhunderts wie ein Wirbelwind durch Europa fegte: den “Sidereus Nuncius” von niemand geringerem als Galileo Galilei. Dieses Buch markierte den Beginn einer Revolution, denn es war eines der ersten wissenschaftlichen Werke, das nicht nur auf Latein, sondern auch in der Volkssprache Italienisch verfasst wurde und so einer breiteren Öffentlichkeit zugänglich war.

Ich habe jetzt eine Variante von Chat-GPT aufgebaut. Für diejenigen unter Ihnen, die bereits mit Chat-GPT gearbeitet haben, wird die Oberfläche vertraut aussehen.

## 47.6 Der Kern der Vorlesung

Doch lassen Sie uns zum Kern dieser Vorlesung kommen. Mir ist es wichtig, dass wir gemeinsam verstehen, was KI eigentlich ist und wie sie funktioniert. # Begrüßung und Einführung in Interaktion mit ChatGPT

Stellen Sie sich vor, Sie öffnen die Seite von ChatGPT und werden mit der freundlichen, typisch amerikanischen Begrüßungsfloskel “How can I help you?” empfangen. Es klingt wesentlich servicefreundlicher als ein schlichtes “Hi, hier bin ich”. Die KI bietet Ihnen direkt ihre Hilfe an und präsentiert vier mögliche Optionen, die zwar oft irrelevant sind, Ihnen aber die Mühe ersparen sollen, sich selbst etwas auszudenken. Darunter können Sie dann eingeben, wobei Sie Unterstützung benötigen.

# **48 Demonstrieren der Möglichkeiten von ChatGPT anhand eines Beispiels**

## **48.1 Übertragen eines Bildes in maschinenlesbaren Text**

Nehmen wir an, Sie haben eine Seite mit komplexen Inhalten vor sich, mit denen Sie in ihrer jetzigen Form nichts anfangen können. Hier kommt die KI ins Spiel: Sie können einfach einen Screenshot der Seite machen und diesen in den Chat-GPT hochladen. Anschließend instruieren Sie die KI mit einer Anweisung wie “Transkribiere das Bild” - und schon erhalten Sie eine nahezu fehlerfreie Übertragung des nicht gerade einfachen Textes in getippte Buchstaben. Eine Leistung, die bis heute kein anderes Programm in dieser Qualität vollbringen kann.

## **48.2 Übersetzen des Textes in eine andere Sprache**

Doch das ist erst der Anfang. Nehmen wir an, Sie verstehen kein Latein - kein Problem. Tippen Sie einfach “Übersetze diesen Text ins Deutsche” ein und schon erhalten Sie eine verständliche, wenn auch noch etwas gewöhnungsbedürftige Übersetzung. Mit ein wenig Feinschliff oder dem Wechsel des Modells lässt sich daraus ein publikationsreifer deutscher Text erstellen. Und das Ganze funktioniert nicht nur für Deutsch und Englisch, sondern für über 150 Sprachen weltweit, darunter auch Japanisch und Koreanisch. Selbst obskure mittelalterliche Quellen stellen kein Hindernis dar.

# **49 Erweiterung der Möglichkeiten durch Phantasie und gezielte Fragestellungen**

Doch jetzt fängt der eigentliche Spaß erst an. Mit dem nun zugänglichen Text eröffnen sich ganz neue Möglichkeiten jenseits der typischen Google-Fragen wie "Wer war Galilei?" oder "Wann lebte er?". Stattdessen können Sie die KI mit Fragen herausfordern, die Google unmöglich beantworten kann. Zum Beispiel: "In welcher Stadt trank Galilei im Mai 1615 ein Glas Wein?". Das Problem liegt hier nicht nur darin, dass Google dieses spezifische Ereignis nicht kennt, sondern dass eine einfache Stichwortsuche prinzipiell nicht ausreicht, um die Antwort zu finden.

## **49.1 Analogie zu Sherlock Holmes**

Stellen Sie sich die KI als eine Art elektronischen Sherlock Holmes vor. Sie nimmt das gesamte Universum an Dokumenten über Galilei zur Kenntnis - seine Briefe, seine historischen Lebensumstände, seine typischen Aktivitäten im Frühjahr 1605. Aus diesen Informationen zieht sie dann Rückschlüsse und generiert eine fundierte Hypothese darüber, wo und wann Galilei wahrscheinlich sein Glas Wein genossen hat. Zwar nicht mit absoluter Sicherheit, aber basierend auf seinen regelmäßigen Lebensumständen. Solche Fragen werden die KI-Modelle in naher Zukunft beantworten können.

## **49.2 Vielfältige Analysemöglichkeiten von Texten**

Doch damit nicht genug. Sie können die KI auch anweisen, eine Tabelle mit allen Verben des Textes zu erstellen oder gezielt nach Verben zu suchen, die ein Lob, eine Ankündigung oder ein Versprechen ausdrücken - selbst wenn Sie die genaue Formulierung nicht kennen. Die Möglichkeiten sind schier grenzenlos.

Ein konkretes Beispiel: Fragen wir die KI, wer sich laut dem Text bewegt. Nach kurzer Bedenkzeit liefert sie die korrekte Antwort: Die vier Planeten bewegen sich zu verschiedenen Zeiten und mit erstaunlicher Geschwindigkeit um den Stern Jupiter - eine Entdeckung, die Galilei machte und die tatsächlich im lateinischen Originaltext erwähnt wird.

# **50 Philosophie als Grundlage für die Möglichkeiten der KI**

Doch wie ist das alles möglich? Die Antwort liegt in der Philosophie - nicht in der Technik. Natürlich brauchen wir auch die technische Infrastruktur, so wie wir Beamer und Notebooks benötigen. Aber der eigentliche Schlüssel zu den Fähigkeiten der KI ist philosophischer Natur. Das wird oft übersehen, doch ich möchte Ihnen zeigen, warum Philosophie hier so entscheidend ist.

## **50.1 Beantwortung von Fragen über Mikrofoneingabe**

Um das Potenzial der KI weiter zu verdeutlichen, können wir auch das Mikrofon aktivieren und eine Frage stellen: "Hat Galilei diese Entdeckung selbst durch Beobachtungen gemacht?". Das System denkt kurz nach und liefert dann die zutreffende Antwort: Ja, laut den Angaben im Text hat Galilei die Entdeckung tatsächlich selbst durch Beobachtungen gemacht.

Das Erstaunliche daran ist nicht nur, dass überhaupt eine Antwort generiert wird, sondern vor allem die Qualität dieser Antwort - trotz Versprechern und spontaner Formulierung mein-erseits.# Einführung in die sprachliche Dimension der KI

Meine Damen und Herren, heute möchte ich Ihnen eine faszinierende und zugleich beunruhigende Entwicklung in der Welt der künstlichen Intelligenz näherbringen. Es geht um die Fähigkeit von KI-Systemen, nicht nur Informationen aus autoritativen Quellen zu sammeln, sondern eigenständig Antworten zu generieren und Inhalte zu erstellen. Diese Entwicklung hat weitreichende Konsequenzen für unser Verständnis von Wissen und Informationsverarbeitung.

## **50.2 Die Möglichkeiten der KI**

Die Möglichkeiten der KI sind atemberaubend und erweitern sich täglich. Lassen Sie mich Ihnen einige Beispiele nennen:

- Übersetzung: KI-Systeme können Texte von einer Sprache in eine andere übersetzen, und zwar mit einer Genauigkeit und Geschwindigkeit, die menschliche Übersetzer in den Schatten stellt.

- Bild-zu-Text-Konvertierung: KI kann Bilder analysieren und deren Inhalt in Textform beschreiben. Dies eröffnet völlig neue Möglichkeiten der Bildverarbeitung und -archivierung.
- Audio-zu-Text-Konvertierung: Gesprochene Sprache kann von KI-Systemen in Echtzeit transkribiert werden, was die Erstellung von Protokollen und Untertiteln erleichtert.
- Textzusammenfassung: Geben Sie der KI ein ganzes Buch, und sie wird Ihnen eine prägnante Zusammenfassung liefern. Dies kann die Recherche und das Studium enorm beschleunigen.
- Text-zu-Audio-Konvertierung: Umgekehrt kann KI auch geschriebenen Text in gesprochene Sprache umwandeln, was neue Möglichkeiten für Hörbücher und Sprachassistenten eröffnet.
- Text-zu-Video-Konvertierung: Hier wird es geradezu unheimlich. KI kann aus Textbeschreibungen realistische Videos generieren, die kaum noch von echten Aufnahmen zu unterscheiden sind.

## 50.3 Die Gefahren der KI

So faszinierend diese Möglichkeiten auch sind, sie bergen auch erhebliche Risiken. Ein zentrales Problem ist das Phänomen der “Halluzination”. Dabei generiert die KI scheinbar plausible Informationen, die jedoch nicht der Realität entsprechen.

Ein Beispiel: Ich fragte eine KI nach dem Namen der zweiten Frau des Mathematikers Leonhard Euler. Die Antwort klang überzeugend, inklusive eines Verweises auf eine Publikation der Petersburger Akademieschriften von 1784. Doch diese Publikation existiert gar nicht, und die genannte Person war nie mit Euler verheiratet.

Solche Halluzinationen können fatale Folgen haben, wenn sie unerkannt bleiben. Wer eine solche Information zitiert, disqualifiziert sich wissenschaftlich für immer. Dieses Problem trat auch bei der Mars-Mission der NASA auf, als eine KI falsche Informationen über einen Erkundungssatelliten verbreitete.

## 50.4 Der sprachliche Kern der KI

Bei all diesen Anwendungen, sei es Bild-, Audio- oder Videoverarbeitung, bildet die Sprache den Kern der KI-Technologie. Selbst bei der Bildanalyse übersetzt die KI zunächst das Bild in eine verbale Beschreibung, bevor sie weiterverarbeitet wird.

Diese Erkenntnis ist philosophisch bedeutsam und erinnert an Wittgensteins These von der Unhintergehrbarkeit der Sprache. Die sprachliche Verbalisierung von Inhalten ist der Dreh- und Angelpunkt der KI, und genau darum soll es in dieser Vorlesung gehen.

Ich werde mich nicht auf die technischen Details der KI-Entwicklung konzentrieren, sondern auf den Umgang mit Sprache in KI-Modellen. Die anderen Medien sind zwar faszinierend, aber letztlich sekundär. Unser roter Faden wird die philosophische Dimension der sprachlichen Verarbeitung in der KI sein.<sup>#</sup> Gefahren und Probleme der künstlichen Intelligenz

Meine Damen und Herren, lassen Sie uns heute über die Schattenseiten der künstlichen Intelligenz sprechen. Wir haben bereits die atemberaubenden Möglichkeiten dieser Technologie gesehen, doch nun ist es an der Zeit, auch die Probleme und Gefahren zu beleuchten, die damit einhergehen.

## 50.5 Das Problem der Halluzinationen

Eines der ersten Probleme, auf das wir stoßen, sind die sogenannten Halluzinationen der KI-Modelle. Ein eindrucksvolles Beispiel dafür lieferte das Supermodell von Google, das auf die Frage "Wer fliegt denn da?" eine Antwort gab, die zwar plausibel klang, aber rein fiktiv war. Ohne Zugriff auf aktuelle NASA-Informationen oder Tagesnachrichten erfand das Modell kurzerhand einen Satellitennamen. Innerhalb einer halben Stunde wurde es vom Netz genommen, und der Marktwert von Google-Aktien sank um Millionen. Seitdem trauen sich die Unternehmen nicht mehr, ihre Modelle zu veröffentlichen.

Doch warum halluzinieren die Modelle überhaupt, wenn sie doch schon so viele Fähigkeiten besitzen? Die Antwort darauf ist komplexer als man denkt.

## 50.6 Die Gefahr der Manipulation durch glaubwürdige Fakes

Ein weiteres Problem, das eng mit den Halluzinationen verbunden ist, ist die Fähigkeit der KI, glaubwürdige Texte, Bilder und sogar Videos zu produzieren. Dies öffnet Tür und Tor für falsche oder manipulative Informationen, die auf den ersten Blick echt erscheinen.

Ein aktuelles Beispiel dafür sind die Videos, die im Zusammenhang mit dem Raketenüberfall auf Israel in den sozialen Medien aufgetaucht sind. Sie zeigten panische Einwohner von Tel Aviv, die vor nicht existierenden Einschlägen flohen. Diese Videos wurden absichtlich generiert, um die Öffentlichkeit zu täuschen, und sind für den Betrachter zunächst nicht als Manipulation zu erkennen.

## **50.7 Selektive Informationen und die Pluralität der Hintergründe**

Jede Antwort, die uns ein KI-Modell gibt, basiert auf bestimmten Annahmen und Voraussetzungen. Diese haben jedoch immer auch Alternativen, die möglicherweise nicht besser oder schlechter sind, aber eine Pluralität an Hintergründen darstellen.

Wenn wir eine bestimmte Antwort akzeptieren, akzeptieren wir auch die Voraussetzungen dafür und vernachlässigen die Alternativen. Ein Beispiel dafür ist die Anfrage an ein KI-Modell, ein Porträt eines möglichen Nachfolgers des jetzigen Papstes zu erstellen. Aufgrund der politisch korrekten Voreinstellung des Modells wurde eine farbige Frau im Papstgewand generiert - eine Darstellung, die in der Realität aufgrund der Zusammensetzung des Kardinalskollegiums höchst unwahrscheinlich ist.

Dieses Beispiel verdeutlicht, wie selektive Informationen zu verzerrten Ergebnissen führen können. Es wirft die Frage auf, wie wir mit diesen Problemen umgehen sollen.

## **50.8 Die Unausweichlichkeit der KI-Entwicklung und die Notwendigkeit der Gestaltung**

Eines ist klar: Wir können uns vor diesen Fragen nicht drücken. Die Entwicklung der künstlichen Intelligenz ist unwiderstehlich und unausweichlich. Ab heute werden uns diese Technologien mit all ihren Vor- und Nachteilen zunehmend beschäftigen.

Wir müssen lernen, damit umzugehen und die Entwicklung aktiv mitzugestalten. Nicht im Sinne einer Kontrolle, sondern einer Gestaltung. Denn wenn wir jetzt nicht eingreifen, laufen wir Gefahr, die Kontrolle über diesen Prozess zu verlieren.

## **50.9 Weitere Gefahren: Diskriminierung und Überwachung**

Neben der selektiven Information gibt es weitere Gefahren, die wir im Auge behalten müssen. Dazu gehören Dimensionen der Diskriminierung, bei denen bestimmte Personengruppen oder Qualifikationen berücksichtigt werden, andere hingegen nicht.

Auch die Möglichkeiten der Überwachung durch KI-Systeme sind alarmierend. Ein Beispiel dafür ist China, wo Besucher bei der Einreise lediglich in eine Kamera lächeln müssen und dann während ihres gesamten Aufenthalts live verfolgt und protokolliert werden.

Diese Entwicklungen werfen Fragen auf, wie weit solche Technologien zugelassen und kontrolliert werden sollten. Eine Antwort darauf zu finden, ist keine leichte Aufgabe.

## 50.10 Die Notwendigkeit der Auseinandersetzung mit KI

Angesichts dieser erschütternden Probleme könnte man geneigt sein, das Thema KI einfach zu vergessen. Wozu sich mit Übersetzungen von Galileis lateinischen Texten beschäftigen, wenn wir dafür doch unsere Gelehrten haben?

Doch so einfach ist es nicht. Die Vorteile der künstlichen Intelligenz sind zu groß, um sie zu ignorieren. Wir müssen uns mit dieser Technologie auseinandersetzen, ihre Möglichkeiten nutzen und gleichzeitig ihre Schattenseiten im Blick behalten. Nur so können wir eine Zukunft gestalten, in der die KI zum Wohle der Menschheit eingesetzt wird. # Begrüßung und Einführung

Einen schönen guten Tag, meine Damen und Herren. Heute möchte ich mit Ihnen über zwei Fragen sprechen, die mir in letzter Zeit immer wieder begegnen. Zunächst einmal habe ich eine Frage zu der Konferenz, von der ich gehört habe, dass sie in diesem Monat August stattfinden soll. Wo genau findet diese Konferenz statt?

Ah, ich verstehe. Es handelt sich also um eine regelmäßig wiederkehrende Konferenzserie, die im August abgehalten wird. Es ist bemerkenswert, wie weit fortgeschritten die Aufmerksamkeit und das Wissen um diese Themen inzwischen auch in den Institutionen sind. Sogar auf EU-Ebene wurde im März letzten Jahres bereits ein Bericht veröffentlicht, in dem diese Angelegenheiten thematisiert wurden. Allerdings wurden sie dort so behandelt, wie ich es gerade geschildert habe - sie wurden angesprochen, aber nicht gelöst.

Es gibt keine einzelne Konferenz, die sich des Problems annimmt und von der wir erwarten könnten, dass es in kürzester Zeit gelöst wird. Nein, so einfach ist es leider nicht. Stattdessen möchte ich auf die philosophischen Aspekte eingehen, die sowohl den Vorteilen als auch den Gefahren zugrunde liegen.

## **51 Beispiele für die Nutzung von Sprachen in der Wissenschaft**

Lassen Sie uns ein Beispiel betrachten, das wir gerade schon diskutiert haben. In der Fachliteratur hält sich hartnäckig das Gerücht, dass Galileis Vater sich negativ über die wissenschaftliche Nutzung anderer Sprachen als Latein geäußert haben soll. Das würde natürlich einen spannenden Vater-Sohn-Konflikt darstellen, denn Galilei selbst ist ja berühmt dafür, dass er das Italienische für die Wissenschaft nutzbar machte, indem er auf Italienisch publizierte.

In zahlreichen Sekundärquellen findet man die These, dass sein Vater dies nicht für wissenschaftlich hielt und dass sein Sohn Galileo Galilei sich besser von diesen italienischen Publikationen fernhalten sollte. Oh, Moment mal - da steht, dass Kepler sich gegenüber Galilei negativ geäußert hat, nicht Galileis Vater. Danke für den Hinweis! Das ist keine Halluzination, sondern ein echter Fehler meinerseits. Ich hoffe, ich vergesse nicht, das für die Internetversion zu korrigieren.

Die Pointe ist jedenfalls, dass man eine solche Frage - ob sich eine Person X irgendwo negativ zu einer bestimmten These geäußert hat - mit Google nicht beantworten kann. Das mag trivial klingen, aber im Moment ist es tatsächlich nicht möglich, dies durch eine Google-Suche herauszufinden. Warum? Weil Google Ihnen kein Dokument im Internet liefern wird, in dem diese Frage direkt beantwortet wird. Und wenn es ein solches Dokument nicht gibt, ist die Frage für Sie mit Google-Techniken nicht zu beantworten.

Dabei handelt es sich um eine Frage, die historisch gesehen entweder wahr oder falsch ist. Wie kann man das also entscheiden? Nicht mit den heutigen Google-Techniken. Hier braucht es eine neue Dimension der Recherche, die über bestimmte Fähigkeiten verfügen muss.

## **52 Aufgaben und Fragen, die mit herkömmlichen Methoden nicht lösbar sind**

Lassen Sie mich Ihnen anhand einer Liste von Aufgaben und Fragen veranschaulichen, wie zunehmend Probleme auftauchen, die mit den heutigen akademischen Techniken nicht zu lösen sind. Ich spreche hier von Fragen, die selbst Sie als forschende Person nicht beantworten können, wenn sie halbwegs komplex sind.

Mir geht es um die unlösbaren Probleme der realen Forschungswelt, die zwar mit KI lösbar wären, aber aufgrund bestimmter fehlender Fertigkeiten bisher nicht gelöst werden können. Jetzt befinden wir uns im philosophischen Teil meiner Ausführungen und ich werde versuchen, dies sprachanalytisch zu komprimieren.

### **52.1 Frage 1: Einfache Aussage in einer Quelle**

Angenommen, Person A äußert sich in einer Quelle Q zu einer Person namens Jochen Schmidt. Ist diese Aussage wahr oder falsch? Hier haben Sie noch eine gewisse Chance, die Frage eindeutig zu beantworten, wenn Sie die Quelle Q gefunden haben und darin die Person A benannt wird und sich zu Jochen Schmidt äußert. Der Anforderungsgrad ist hier noch nicht sehr hoch. Wenn das Ihre Examensaufgabe wäre, hätten Sie eine realistische Chance, sie zu lösen. Sie müssten nur so lange alle Quellen durchlesen, bis Sie die richtige gefunden haben.

### **52.2 Frage 2: Aussage in Briefen zu einem Thema**

Nehmen wir an, Person A äußert sich in ihren Briefen zu einem Thema T. Das können Sie schon nicht mehr ohne weiteres lösen, ohne eine Lebensdauer damit zu verbringen, das gesamte Schrifttum von Person A zu lesen. Wenn Sie z.B. für eine Examsarbeit eine Biografie über eine Person namens Heinz Müller verfassen sollten und eine solche Aufgabe hätten, müssten Sie zunächst alle Briefe zusammentragen und sie komplett lesen. Und selbst dann wären Sie sich nicht sicher, ob Sie wirklich alle Briefe gefunden haben.

Denken Sie nur an die Kafka-Forscher. Wenn Sie wissen wollen, ob sich Kafka in seinen Briefen jemals zu einem bestimmten Thema geäußert hat oder nicht, haben Sie einen enormen manuellen Forschungsaufwand vor sich, um überhaupt in die Nähe einer Antwort zu kommen.

Hier befinden wir uns bereits in Bereichen, die schwer zu beantworten sind - Fragestellungen, die bislang praktisch nicht zu lösen waren.

### **52.3 Frage 3: Aussagen einer Person in ihren Schriften**

Hat eine Person A in ihren Schriften Aussagen der Art T getroffen, wenn Person A sehr viel geschrieben hat? Nehmen wir als Beispiel die Briefe Napoleons. Hat sich Napoleon jemals zu Aspekten der Vorläufer der Genfer Konvention bei der Kriegsführung geäußert? Das können Sie aus praktischen Gründen nicht lösen. Ich will an dieser Stelle nicht sagen, dass es prinzipiell unmöglich ist, aber in der Wissenschaft möchte man solche Fragen beantwortet haben. Und das gilt nicht nur für das öffentliche Interesse, sondern auch für die Wissenschaft selbst.

Sie können sich vorstellen, Welch enorme Konsequenzen es für die Wissenschaft hätte, wenn man solche Fragen überhaupt beantworten könnte. Dann wäre es möglich, weitreichende Thesen zu Napoleons Verständnis von Krieg und Frieden aufzustellen, die von der Evidenz abhängen, mit der man solche Fragen beantworten kann. Im Moment ist das nicht möglich.

### **52.4 Frage 4: Keine Aussage einer Person in ihren Schriften**

Angenommen, Person A hat in ihren Schriften keine Aussage T getroffen. Als normaler arbeitender Historiker oder Geisteswissenschaftler werden Sie diese Frage nicht seriös beantworten können. Deshalb gibt es in der Literatur die Unsitte, andere Werke zu zitieren, die sich aus irgendwelchen Gründen dazu bemüßt fühlten, solche Fragen zu beantworten.

Ein Beispiel: Nehmen wir wieder Kafka. Manche Autoren vertreten die These, dass Kafka sich nie antisemitisch geäußert hat. Aber welche Evidenz können Sie dafür eigentlich angeben? Es ist schwierig, eine nicht vorhandene Lektüre von Briefen als Beleg anzuführen. Wie wollen Sie eine solche These rechtfertigen, wenn Sie sie vertreten?

Eine der größten Unsitten der gegenwärtigen akademischen Literatur besteht darin, nicht selbst das Risiko einer These einzugehen, sondern stattdessen den berühmten Heinz Müller zu zitieren, weil er schon einmal etwas Ähnliches gesagt hat. Also fügt man eine Fußnote in die Arbeit ein: "Heinz Müller, 1973, Seite 5: Ganz klar, Kafka hat sich nie antisemitisch geäußert." Und auf einmal entsteht ein Schneeballsystem, das dem Halluzinationseffekt ähnelt, den wir gerade hier hatten. Und zwar nur deshalb, weil die Evidenz, die für bestimmte Thesen erforderlich ist, auf manuelle Weise kaum zu beschaffen ist. Mit KI werden Sie das in Zukunft können.

# **53 Die Herausforderung der inhaltlichen Analyse mit KI**

Jetzt werden Sie vielleicht fragen: Inwiefern ist das speziell für KI relevant? Man könnte doch erwarten, dass sich das grammatisch lösen lässt. Wenn ich die Aussage T formalisieren kann, müsste ich doch auf dem Textkorpus einfach prüfen können, ob diese Bedingung irgendwo erfüllt ist, oder?

Genau das ist der springende Punkt, und ich muss jetzt ein bisschen auf die Uhr schauen, damit ich meine Kurve hier noch hinbekomme. Aber diese Kurve berührt schon das Thema. Was heißt es, in Ihrem Korpus prüfen zu können?

Nehmen wir an, Sie hätten den Idealfall: Kafkas gesammelten Briefwechsel in einer Datenbank. Jetzt möchten Sie wissen, ob es darin eine antisemitische Formulierung gibt. Wie sieht die denn aus? Wenn Sie Ihre Datenbank nach Art einer Google-Suche nach bestimmten Wortvorkommnissen durchforsten, dann können Sie das lösen. Das ist die klassische Vorgehensweise.

Aber inhaltlich betrachtet: Was ist eigentlich eine antisemitische Äußerung? Sobald es darum geht - und deshalb habe ich es hier erwähnt - kön# Betrachtungen zur künstlichen Intelligenz und Sprachverarbeitung

Meine sehr geehrten Damen und Herren, liebe Studierende,

in der heutigen Vorlesung möchte ich Ihnen einen faszinierenden Einblick in die Welt der künstlichen Intelligenz und insbesondere deren Fähigkeiten zur Sprachverarbeitung geben. Wir werden uns mit der Frage beschäftigen, inwieweit KI-Systeme in der Lage sind, komplexe sprachliche Konstrukte wie Metaphern, Ironie oder versteckte Bedeutungen zu erkennen und zu interpretieren.

## **53.1 Grenzen der traditionellen Datenbanken**

Zunächst einmal möchte ich klarstellen, dass ich keineswegs behauptet habe, es gäbe in den vorliegenden Dokumenten keine relevanten Satzvorkommnisse. Die herkömmliche Art der Dokumentenaufzeichnung und -abfrage, wie sie etwa mit Datenbanken möglich ist, erlaubt zwar das Auffinden bestimmter Textpassagen, jedoch keine inhaltlichen Suchen im eigentlichen Sinne.

Selbst moderne KI-Systeme können nicht mit absoluter Sicherheit feststellen, dass eine bestimmte Aussage nicht getroffen wurde, da stets die Möglichkeit besteht, dass die zugrunde

liegende Datenbasis unvollständig ist. Vielmehr lässt sich hier nur mit Wahrscheinlichkeiten operieren - ein Begriff, den ich an dieser Stelle allerdings kritisch hinterfragen möchte.

## **53.2 Qualifizierte Aussagen auf Basis der verfügbaren Evidenz**

Wahrscheinlichkeiten sind numerische Werte zwischen 0 und 1, die man in diesem Kontext nicht sinnvoll einsetzen kann. Stattdessen sollte man sich auf die konkrete Situation beziehen und feststellen: Auf Basis dieser und jener Grundgesamtheit von Briefwechseln und Äußerungen, die als Dokumente für die Befunde zur Verfügung stehen, lässt sich unter der Voraussetzung, dass sie die alleinige Entscheidungsgrundlage bilden, folgendes Fazit ableiten.

Eine solche differenzierte Betrachtung der Befundlage ist unerlässlich, denn es lässt sich ja nicht ausschließen, dass genau jene Briefe, die möglicherweise relevante Inhalte enthalten, vernichtet wurden. Ein solches Szenario würde den Wahrheitswert der Fragestellung grundlegend verändern. Auch KI-Systeme können diese Problematik nicht vollständig ausräumen, sehr wohl aber eine qualifizierte, auf der verfügbaren Evidenz basierende Antwort geben.

## **53.3 Herausforderungen bei der Interpretation von Metaphern und Ironie**

Ein besonders spannendes Feld ist die Fähigkeit von KI-Systemen, mit Metaphern und ungewöhnlichem Sprachgebrauch umzugehen. Gerade im Kontext des Antisemitismus verbergen sich oft codierte Botschaften hinter scheinbar harmlosen Formulierungen. Während eine Blut- und Boden-Ideologie relativ leicht zu identifizieren ist, stellt die Interpretation von Begriffen wie "entwurzelt" oder "ohne Verwurzelung" eine ungleich größere Herausforderung dar.

Anhand eines konkreten Beispiels möchte ich Ihnen verdeutlichen, wozu moderne KI-Systeme in diesem Bereich bereits in der Lage sind. In München hatten wir es mit revolutionären Briefen aus der Zeit der Französischen Revolution zu tun, die in elegantem Französisch verfasst waren und vor Ironie und Sarkasmus nur so strotzten. Um diese Feinheiten zu erkennen, bedarf es zunächst einmal exzellenter Sprachkenntnisse. Doch selbst dann gilt es, die ironischen Komponenten als solche zu identifizieren.

Ich kann Ihnen versichern, dass KI-Systeme mittlerweile über eine Sprachkompetenz verfügen, die es ihnen erlaubt, auch diese Dimension der Sprachverwendung zu erkennen. Allerdings dürfen Sie sich das nicht als simples Schwarz-Weiß-Schema vorstellen, bei dem man einfach einen "Ironie-Kompetenz-Knopf" umlegt und schon funktioniert alles wie bei einem literarischen Meisterinterpreten.

## **53.4 Lernfähigkeit und Entwicklungspotenzial von KI-Systemen**

Vielmehr müssen Sie sich den Lernprozess der KI ähnlich vorstellen wie Ihre eigene Entwicklung zu Beginn Ihres Studiums. Auch Sie haben im Laufe der Zeit eine Menge dazugelernt und sich weiterentwickelt. Genauso können auch KI-Modelle lernen und sich verbessern. Ich möchte keineswegs behaupten, dass bereits alle Probleme und Herausforderungen gelöst sind, aber es gibt vielversprechende Lösungsansätze, um auch mit komplexeren Formen der Sprachverwendung umgehen zu können.

In München haben wir beispielsweise erfolgreich getestet, ob KI-Systeme in der Lage sind, bissige Karikaturen aus den 1920er Jahren zu interpretieren und zu erkennen, welche Personen mit welchen Klischees auf den Arm genommen werden. Mit dem richtigen Training ist es den Bilderkennungsalgorithmen tatsächlich gelungen, diese Zusammenhänge zu entschlüsseln.

## **53.5 Der Paradigmenwechsel durch Large Language Models und Embeddings**

Der entscheidende Unterschied und gleichzeitig der Punkt, an dem der “Philosophical Turn” der KI einsetzt, liegt in der Entwicklung von Techniken wie Large Language Models oder Embeddings. Diese ermöglichen eine Abkehr von der reinen Textsuche hin zu einer inhaltlichen Erfassung der Bedeutung sprachlicher Ausdrücke. Dieser semantische Wechsel, den ich auch gerne als “Semantic Turn” bezeichne, ist der Schlüssel zu den beeindruckenden Fähigkeiten moderner KI-Systeme.

Egal ob es um die Analyse von Bildern, Texten oder Audioaufnahmen geht - all diesen Anwendungen liegt zugrunde, dass die Systeme nicht nur nach bestimmten Zeichenfolgen suchen, sondern deren Bedeutung erfassen und identifizieren können. Genau darum geht es bei den milliardenschweren Investitionen in diesem Bereich: den Modellen beizubringen, auf Basis der eingegebenen Daten die dahinterstehende Semantik zu erkennen.

## **53.6 Die Bedeutung der Philosophie für die KI-Forschung**

Damit eröffnet sich ein weites Feld für die Philosophie. Solange wir nur von Sätzen sprechen, bewegen wir uns auf der Ebene von Formulierungen und syntaktischen Strukturen. Wenn wir jedoch nach der Bedeutung eines Ausdrucks fragen, betreten wir Neuland. Genau hier setzt die aktuelle KI-Revolution an, und deshalb ist die Philosophie von zentraler Bedeutung für diese Entwicklung.

Als Studierende der Philosophie sollten Sie mit der klassischen Unterscheidung zwischen Satz und Aussage vertraut sein. Im Deutschen ist diese Differenzierung von größter Wichtigkeit, während sie in englischen Übersetzungen oft vernachlässigt wird. So haben etwa die Übersetzer

von Wittgensteins Gesammelten Werken sowohl für “Aussage” als auch für “Satz” durchgängig den Begriff “Sentence” verwendet, was zu erheblichen Missverständnissen führen kann. Im Englischen heißt es korrekterweise “Sentence” für Satz und “Proposition” für Aussage.

Genau diese Unterscheidung markiert die fundamentale Revolution, die sich gerade vollzieht: Wir haben es nun mit Maschinen zu tun, die mit Aussagen umgehen können. Und nur Aussagen, nicht Sätze, können wahr oder falsch sein. Wer also über Fake News, Halluzinationen und ähnliche Phänomene spricht und sich dabei auf Sätze bezieht, liegt philosophisch gesehen völlig falsch. Wahrheit und Falschheit können sich konzeptionell nur auf Aussagen beziehen.

Die Tatsache, dass KI-Systeme nun in der Lage sind, sich mit Aussagen zu befassen, birgt ebenso faszinierende Möglichkeiten wie Gefahren. In der nächsten Vorlesung werden wir uns eingehender mit diesen Aspekten beschäftigen und uns ansehen, wie genau diese neuen Technologien funktionieren und welche Auswirkungen sie haben können.

Ich danke Ihnen für Ihre Aufmerksamkeit und freue mich darauf, dieses spannende Thema in der kommenden Woche gemeinsam mit Ihnen zu vertiefen. # Begrüßung zur zweiten Vorlesung Philosophie der AI

Herzlich willkommen, meine Damen und Herren, zur zweiten Vorlesung unserer Reihe “Philosophie der AI”. Lassen Sie uns heute an die spannenden Erkenntnisse der letzten Sitzung anknüpfen und gemeinsam ergründen, welche faszinierenden Möglichkeiten die Künstliche Intelligenz für die geisteswissenschaftliche Forschung bereithält. Stellen Sie sich vor, wie AI unsere alltägliche Arbeit nicht nur erleichtern, sondern revolutionieren und bisher ungeahnte Perspektiven eröffnen kann.

# **54 Traditionell schwer lösbarer Fragen in der Forschung**

In der Welt der Wissenschaft gibt es eine Vielzahl von Fragestellungen, die uns immer wieder vor Herausforderungen stellen und deren Beantwortung mit herkömmlichen Mitteln oft an Grenzen stößt. Nehmen wir beispielsweise die Suche nach Evidenz in einem definierten Kreis von Quellen, einem sogenannten Scholarium, um eine historische Aussage H zu belegen. Jeder von Ihnen, der schon einmal eine wissenschaftliche Arbeit verfasst hat, weiß, wie zeitaufwändig und mühsam dieser Prozess sein kann - je nach Komplexität der Fragestellung. Doch mit der Unterstützung von AI könnten wir in Zukunft, abhängig von der Zugänglichkeit und Aufbereitung des Scholariums, solche Nachweise schnell und effizient führen.

## **54.1 Evidenz finden, um eine Hypothese zu widerlegen**

Noch kniffliger wird es, wenn wir in einem Scholarium nach Evidenz suchen, um eine Hypothese H zu widerlegen. Im Alltag des Wissenschaftlers ist dies praktisch unmöglich - und dennoch finden wir solche Aussagen häufig in Publikationen. Überlegen Sie selbst: Wie oft haben Sie schon in Hausarbeiten, Qualifikationsschriften oder Fachartikeln Behauptungen gelesen, die eine These anhand von Standardreferenzliteratur zu widerlegen versuchen? Häufig sucht man vergeblich nach der tatsächlichen Evidenz dafür. Stattdessen wird allzu oft der bequeme Weg gewählt, sich auf Kollegen zu berufen, die ähnliche Aussagen getroffen haben - doch das ist keine echte Evidenz, sondern bestenfalls eine fragwürdige Praxis.

## **54.2 Zeitgenössische Autoren und ihre Äußerungen zu historischen Hypothesen**

Lassen Sie uns noch einen Schritt weiter gehen und uns einer noch komplexeren Fragestellung zuwenden: Welcher zeitgenössische Autor hat sich zu einer historischen Hypothese H ebenfalls geäußert? Stellen Sie sich vor, Sie interessieren sich für eine bestimmte These, die der wissenschaftshistorische Autor Johannes Kepler im Jahre 1603 aufgestellt hat. Nun möchten Sie wissen, welche seiner Zeitgenossen sich zu ähnlichen Fragen geäußert haben. Ohne jahrelange akribische Lektüre und Archivarbeit ist eine solche Recherche praktisch unmöglich.

### **54.3 Der Einfluss von Publikationen auf historische Autoren**

In wissenschaftlichen Feststellungen stoßen wir oft auf Aussagen wie: "Wer hat die Publikation von H, eines historischen Autors, relevant beeinflusst?" Doch Hand aufs Herz - die meisten dieser Behauptungen sind spekulativ und unbegründet. Nicht etwa, weil die Forscher unseriös arbeiten, sondern weil der Evidenznachweis für solche Aussagen extrem schwierig zu führen ist. Schon allein die Frage, was genau einen "relevanten Einfluss" auf eine Hypothese ausmacht, ist alles andere als trivial.

# **55 Die Rolle der AI in der geisteswissenschaftlichen Forschung**

Hier zeigt sich deutlich, dass die Diskussion um AI weit über eine rein technische Erleichterung unserer Arbeit hinausgeht. Vielmehr eröffnet sie uns die Möglichkeit, alltägliche Fragestellungen überhaupt erst bearbeitbar zu machen, die bislang nur unzureichend gelöst werden konnten. Nehmen wir als weiteres Beispiel die Frage, wer eine Alternative zu einer historischen Hypothese H vertreten hat. Schon die Definition dessen, was eine "Alternative" in diesem Kontext bedeutet, ist eine Herausforderung - von der Suche nach entsprechenden Äußerungen in der Gesamtliteratur eines Scholariums ganz zu schweigen. Praktisch unmöglich, wenn auch theoretisch denkbar.

## **55.1 Die Gefahren der AI und ihre Korrektur durch verbesserte Praktiken**

In der letzten Vorlesung haben wir unter der Rubrik "Gefahren der AI" diskutiert, wie Aussagen oder Befunde, die mittels AI generiert wurden, selektiv sein können, halluzinierte Thesen vertreten oder auf andere Weise kritisch hinterfragt werden müssen. Doch heute betrachten wir die Kehrseite der Medaille: AI kann auch dazu beitragen, unzulängliche oder problematische Praktiken in der gegenwärtigen Forschung zu korrigieren oder gar gänzlich zu ersetzen. Meine These lautet daher unmissverständlich: Der Eingriff von AI in unser wissenschaftliches Tagesgeschäft wird unsere Disziplinen in kürzester Zeit, in wenigen Jahren, drastisch verändern. Mein Rat an Sie: Beschäftigen Sie sich so schnell wie möglich mit diesen Mitteln, auch schon während Ihres Studiums - andernfalls werden viele Fragen Ihrer Qualifikationsarbeiten nicht mehr den zukünftigen Anforderungen genügen.

## **55.2 Nachvollziehbarkeit von Begründungen für historische Hypothesen**

Ein weiteres Beispiel für eine bislang schwer zu beantwortende Frage ist, inwiefern die Begründung für eine historische Hypothese H für andere Zeitgenossen nachvollziehbar oder überzeugend gewesen sein mag. Wenn wir wissenschaftliche Kontroversen einer bestimmten Epoche verstehen wollen, müssen wir uns fragen: Warum konnte die Publikation eines Autors A seine

Kollegen B nicht überzeugen? Ein klassischer Fall in der Wissenschaftsgeschichte ist das Werk “De revolutionibus” von Kopernikus, das zwar einige, aber bei weitem nicht die Mehrheit seiner Zeitgenossen überzeugen konnte. Doch warum war das so? Spekulationen führen uns hier nicht weiter - stattdessen müssen wir solche Fragen auf eine solide methodische Grundlage stellen, und dies ist nur mittels KI möglich.

# **56 Die Entwicklung der KI und ihr Einfluss auf das wissenschaftliche Arbeiten**

Doch wie genau kann die KI uns bei der Beantwortung solch komplexer Fragen unterstützen? Dieser Frage werden wir uns im Laufe der Vorlesung eingehend widmen - und ich bin zuverlässig, dass wir gemeinsam Antworten finden werden. Dabei werden wir feststellen, dass die Anwendung von KI weit weniger komplex und kompliziert ist, als man zunächst vermuten mag. Vielmehr ist sie das Ergebnis eines Jahrzehntelangen Entwicklungsprozesses, der nun zu einem Leistungssprung führt, der auf den ersten Blick wie eine einmalige technische Neuerung aus dem Nichts erscheinen mag. Doch dieser Eindruck täuscht: Tatsächlich handelt es sich um einen langen evolutionären Prozess, der jetzt zu einem qualitativen Umbruch führt.

## **56.1 Die Entwicklung von Interfaces zur Interaktion mit KI**

Diese Entwicklung lässt sich anschaulich an der Art und Weise ablesen, wie wir als Nutzer mit diesen Technologien interagieren. Vor etwa 25 Jahren, genauer gesagt vor eher 20 Jahren, wurde der Browser erfunden - ein technisches Hilfsmittel, mit dem wir aufbereitete HTML-Webseiten betrachten können. Der entscheidende Clou dieser am CERN entwickelten Technik bestand darin, dass die Seiten Verlinkungen zu anderen Seiten enthielten und so ein schnell wachsendes Netzwerk an Informationen bereitstellten. Vor rund 15 Jahren folgte dann die Erfindung des Smartphones, das heute aus unserem Alltag kaum noch wegzudenken ist und uns einen ähnlichen Zugriff auf Inf# Einführung in die Interaktion mit KI-Systemen

Heutzutage interagieren wir hauptsächlich über drei grundlegende Techniken mit künstlicher Intelligenz und den dadurch bereitgestellten Informationen. Die erste und wohl bekannteste Methode ist die Eingabe über ein Textfeld, das vor 25 Jahren mit dem HTTP-Protokoll definiert wurde. Dieses Feld, das oft fälschlicherweise als "Eingabefeld" bezeichnet wird, ermöglicht es Ihnen, mithilfe der Tastatur Links zu anderen Quellen einzugeben. Sie kennen diese Funktion vom Browser, wo Sie in der Adresszeile einen Link eintippen können.

## **56.2 Entwicklung der Eingabemöglichkeiten**

Ursprünglich diente das Adressfeld ausschließlich dazu, Verknüpfungen zu anderen Seiten und Adressen einzugeben. In den letzten 15 Jahren hat sich dieses Feld jedoch weiterentwickelt

und erlaubt nun die Eingabe von weiteren Anfragen – eine Funktion, die technisch gesehen gar nicht so anspruchsvoll ist, aber enorme Auswirkungen hat. Die berühmte Google-Suche ist ein perfektes Beispiel dafür: Anstatt selbst die Webadressen weiterer Quellen eingeben zu müssen, überlassen Sie diese Aufgabe nun einer Suchmaschine, die Ihre Anfrage beliebiger Art verarbeitet und Ihnen die entsprechenden Suchergebnisse zurückgibt.

### 56.3 Der Aufstieg von Chat-GPT

Mit der Einführung von Chat-GPT erleben wir einen massiven Umbruch in der Interaktion zwischen Mensch und Maschine. Chat-GPT, eine dialogorientierte Seite, die die Interaktion mit der KI ermöglicht, hat einen tieferen Grund für ihren Erfolg, den ich gleich noch erläutern werde. Zunächst mag es wie ein cleverer Marketing-Trick erscheinen, doch tatsächlich war es der erste höchst erfolgreiche Auftritt der KI-Modelle über eine solche Chat-Interaktion.

Wir befinden uns derzeit an einem Wendepunkt, an dem sich die Art und Weise, wie wir mit Maschinen interagieren, radikal verändert. Was vorher nur dazu diente, Inhalte von Providern bereitzustellen, wird sich jetzt zu einer Interaktion mit einem KI-Modell entwickeln. Sie werden nicht mehr mit einem Provider kommunizieren, sondern mit einem KI-Modell interagieren, das Ihre Informations- und Mitteilungsbedürfnisse steuert.

### 56.4 Weitere Interaktionsmöglichkeiten

Neben der Texteingabe gibt es noch weitere Möglichkeiten, mit KI-Systemen zu interagieren:

- Sprachbefehle und Spracheingaben, wie Sie sie von Siri kennen, ermöglichen es Ihnen, Befehle über das Mikrofon eines Computers einzugeben, die dann in entsprechende Befehlsstrukturen umgesetzt werden und eine Reaktion der Maschine auslösen.
- Datenbrillen und Headsets eröffnen neue Möglichkeiten der Interaktion. Obwohl ein erster Versuch von Google vor sechs Jahren aufgrund von Bedenken hinsichtlich der Privatsphäre scheiterte, planen nun alle größeren Firmen die Einführung solcher Geräte. Als Tourist könnten Sie beispielsweise vor einem Monument in Rom stehen und über die Brille Informationen zu dessen Erbauung und Geschichte abrufen. Oder Sie sitzen in der Oper und lassen sich eine Szene über den Ohrhörer erläutern.
- Gesten, sowohl taktile als auch sichtbare, können als Signale für die KI dienen. Bei vollständig gelähmten Personen gibt es sogar Implantate, die Hirnströme nutzen, um Signale nach außen zu senden.

Die Entwicklung in all diesen Bereichen schreitet rasant voran, und es bleibt spannend zu beobachten, in welche Richtung sie sich in den nächsten Jahren bewegen wird.

# 57 Die Architektur hinter den KI-Systemen

## 57.1 Generative KI

Auf den ersten Blick mag die Funktionsweise der Software und Programme im Hintergrund höchst kompliziert erscheinen – schließlich können derzeit nur die größten Konzerne mit Milliardenaufwand solche Modelle erstellen. Doch im Kern ist die Architektur gar nicht so kompliziert, wie wir gleich sehen werden.

Es geht hier um die sogenannte generative KI, oft auch abgekürzt als Gen-AI (nicht zu verwechseln mit dem Begriff "Gen"). Diese Systeme erzeugen etwas, das einen bedeutungsvollen sprachlichen Ausdruck darstellt – und genau das ist der revolutionäre Aspekt. Bisher bestanden die Techniken aus Zeichenfolgen, die lediglich eine bestimmte Regelhaftigkeit, eine Syntax, erfüllten, um als bedeutungsvolle Zeichenkette zu erscheinen. Ein Beispiel dafür wäre ein Satz, der im Deutschen durch einen Satzpunkt beendet und durch eine Großschreibung begonnen wird. Andere europäische Sprachen kennzeichnen die Syntax von Sätzen auf unterschiedliche Weise, aber darauf kommt es hier nicht an.

## 57.2 Von der Syntax zur Semantik

Bisher beschränkte sich der Umgang von Computern mit unserer sprachlichen Welt auf die Verarbeitung von Zeichenketten – auf die Syntax. Doch jetzt kommt etwas völlig Neues hinzu, und das ist die große Stunde der Philosophie: die Semantik.

- Die Syntax ist der sprachliche Ausdruck, die Zeichenketten, die Abfolge von Buchstaben, Wörtern und Sätzen, die linear verknüpft werden, um beispielsweise ein Buch zu bilden. All das sind sprachliche Ausdrücke oder, wie es der Philosoph Frege formulierte, der sinnlich wahrnehmbare Ausdruck sprachlicher, gedanklicher Inhalte.
- Die Semantik hingegen befasst sich mit der Bedeutung dieser Zeichen. Und genau damit haben wir es hier zum ersten Mal durch die Technik zu tun.

Im Gegensatz zu den vollmundigen Behauptungen der Konzerne, die schon von "Knowledge Graphen" à la Google sprachen, als von Bedeutung noch gar nicht die Rede war, sollte man diese Terminologie philosophisch hinterfragen. Dann wird schnell klar, dass das Kartenhaus

ziemlich schnell zusammenfällt. Es handelt sich nicht um "Knowledge Graphen", sondern um ganz einfache Graphen. Von Wissen ist da noch keine Spur.

Die philosophische Kritik an der Terminologie entlarvt, was hinter diesen Begrifflichkeiten eigentlich steckt. Und jeder, der bisher von der Bedeutung einer Aussage eines Computers gesprochen hat, weiß nicht, was es üblicherweise in der analytischen Philosophie bedeutet, von Bedeutung zu reden.<sup>#</sup> Die AI-Revolution: Sprache und Bedeutung

Meine Damen und Herren, heute möchte ich Ihnen eine faszinierende Entwicklung näherbringen, die unser Verständnis von Sprache und Bedeutung grundlegend verändern wird: die AI-Revolution. Im Kern geht es darum, dass künstliche Intelligenz nun in der Lage ist, sprachliche Ausdrücke mit ihrer Bedeutung zu verbinden. Diese Fähigkeit hat weitreichende Konsequenzen, die ich Ihnen heute andeuten möchte.

### **57.3 Von der Suche nach Zeichenketten zur Suche nach Inhalten**

Bisher waren Suchmaschinen wie Google darauf beschränkt, nach Zeichenketten zu suchen. Sie gaben einen Begriff ein und die Maschine suchte nach passenden Wörtern, Namen oder Adressen. Damit ließ sich schon viel erreichen, aber im Grunde war es nichts anderes als eine Suche nach Zeichenfolgen.

Doch nun eröffnet sich eine völlig neue Dimension: Die Suche nach den Inhalten und Aussagen, die mit sprachlichen Ausdrücken getroffen werden können. Lassen Sie mich das an einem einfachen Beispiel verdeutlichen:

Der Satz "Der Hund ist schwarz" ist zunächst einmal eine Zeichenkette. Doch diese Zeichenkette ist noch kein Inhalt. In der Philosophie unterscheiden wir streng zwischen dem Satz selbst und der Bedeutung, die er ausdrückt.

### **57.4 Sätze, Aussagen und Wahrheitswerte**

Sätze sind weder wahr noch falsch - eine Aussage, die manchen Informatikern vielleicht überraschend erscheinen mag. Sätze sind sprachliche Ausdrücke, die wohlgeformt sein können, aber keinen Wahrheitswert haben. Wahr oder falsch sind hingegen die mit Sätzen ausgedrückten Inhalte, die wir in der Philosophie als Aussagen oder Propositionen bezeichnen.

Solange wir uns nur auf der Ebene der Syntax bewegen, sind wir noch nicht einmal in der Welt des Wahren und Falschen angelangt. Und ohne Wahrheit oder Falschheit können wir auch nichts glauben oder für richtig halten. Überzeugungen entwickeln wir erst, wenn wir es mit etwas zu tun haben, das wahr oder falsch sein kann - eben mit Aussagen.

## 57.5 Die epistemische Dimension des Wissens

Aussagen sind die Träger von Wahrheitswerten. Und erst wenn wir von Aussagen mit Wahrheitswerten sprechen, kommen wir in die Sphäre der Rechtfertigung, der Kritik und der Widerlegung. Die epistemische Seite des Wissens - das Behaupten, Finden, Kritisieren und Widerlegen von Wissen - setzt voraus, dass wir es mit Aussagen und ihren Wahrheitswerten zu tun haben.

Eine Suchmaschine, die nur Zeichenketten findet, können Sie nicht kritisieren. Sie hat ihre Aufgabe erfüllt, auch wenn das Ergebnis vielleicht nicht Ihren Erwartungen entspricht. Kritik wäre hier fehl am Platz, ja geradezu ein Kategorienfehler.

## 57.6 Die Maschine lernt, Aussagen zu treffen

Wie aber gelingt es nun der Maschine, Aussagen zu treffen - eine Fähigkeit, die wir bisher nur dem menschlichen Geist, der Vernunft zugeschrieben haben? Die AI-Modelle werden derzeit mit höchstem Aufwand darauf trainiert,

1. bedeutungsähnliche Begriffe und Sätze zu unterscheiden,
2. den Strom der sprachlichen Zeichen in Wörter und Satzzeichen zu zerlegen (sogenannte Token),
3. und schließlich die Bedeutungsähnlichkeit von Sätzen zu erkennen.

Nehmen wir drei Beispielsätze:

- An eagle flies silently over the large tree.
- A swan flies noisily over the large tree.
- A mouse eats happily a piece of cheese.

Intuitiv erkennen wir sofort, dass die ersten beiden Sätze in ihrer Bedeutung ähnlich sind, auch wenn die Vögel verschieden sind und sich ihre Art zu fliegen unterscheidet. Wir würden sagen: Es fliegt ein Vogel auf eine bestimmte Weise über einen Baum.

Mit dieser Formulierung greifen wir automatisch auf die Bedeutung der Ausdrücke zu. Wir verallgemeinern soweit, dass wir von Vögeln sprechen, obwohl das Wort "Vögel" gar nicht vorkommt. Wir erkennen die semantische Ähnlichkeit der Sätze.

Der dritte Satz hingegen hat inhaltlich kaum etwas mit den ersten beiden gemeinsam. Allenfalls könnte man sagen, dass es auch hier um ein Tier geht. Aber sonst?

Genau diese Fähigkeit, Bedeutungsähnlichkeiten zu erkennen und Aussagen zu treffen, wird den AI-Modellen nun beigebracht. Und damit eröffnet sich eine völlig neue Dimension der Sprachverarbeitung, die weit über die bloße Suche nach Zeichenketten hinausgeht.<sup>#</sup> Einführung

Meine sehr verehrten Damen und Herren,

lassen Sie uns heute eine faszinierende Reise in die Welt der künstlichen Intelligenz und der Sprachverarbeitung unternehmen. Wir werden ergründen, wie es möglich ist, dass Maschinen die Bedeutung von Wörtern und Sätzen erfassen können - eine Fähigkeit, die lange Zeit als einzigartig menschlich galt.

## 57.7 Die Revolution der Sprachmodelle

Die erste Revolution auf diesem Gebiet ereignete sich, als man begann, die Sprachmodelle mit praktisch der Gesamtheit aller im Internet verfügbaren Texte zu trainieren. Wir sprechen hier von Trillionen von Worteinheiten, sogenannten Token, die als Grundlage dienten. Nicht irgendwelche speziellen Texte, sondern alles, was überhaupt im Netz zu finden ist, wurde herangezogen, um etwas zu definieren, das technisch gesehen als "Embedding" oder "Einbettung" bezeichnet wird.

### 57.7.1 Das Prinzip der Embeddings

Lassen Sie mich das Prinzip der Embeddings näher erläutern. Es handelt sich dabei um komprimierte Zahlenwerte, die Aufschluss darüber geben, in welchem Verwendungszusammenhang bestimmte Wörter mit anderen Wörtern stehen können. Im Grunde genommen werden gigantische Tabellen erstellt, die nichts anderes tun, als zu registrieren, welches Wort welchem anderen Wort folgt und in welchem Kontext von anderen Wörtern es auftritt.

Über mathematisch raffinierte Verfahren, auf die wir hier nicht näher eingehen müssen, lassen sich diese Tabellen so weit kombinieren und komprimieren, dass am Ende eine Tabelle mit 1500 Spalten ausreicht, um jedem einzelnen Satz eine Zuordnung zu geben, welche Rolle jedes Wort innerhalb dieses Satzes für die Bedeutung spielt. Das ist eine erstaunlich geringe Zahl, wenn man bedenkt, wie komplex Sprache ist.

### 57.7.2 Die Bedeutung eines Satzes

Oft wird vereinfachend gesagt, dass dieser Zahlenwert von 1536 Zahlen die Bedeutung eines Satzes ausdrückt. Das ist jedoch nicht ganz korrekt. Zunächst einmal drückt er nur die Kombinationshäufigkeit der Wörter untereinander aus - ein Schritt vor der eigentlichen Frage nach der Bedeutung. Aber es bringt uns der Antwort näher.

## **57.8 Die Herausforderung der Bedeutungsgleichheit**

Eine der ersten Herausforderungen, denen sich die KI-Forschung stellte, war die Frage, welche verschiedenen sprachlichen Ausdrücke die gleiche Bedeutung haben. Eine einfache Frage, die jedoch schwierig zu beantworten ist: Wie kann man maschinell erkennen, dass unterschiedliche Sätze, die von der Syntax her verschieden sind, dennoch das Gleiche ausdrücken?

### **57.8.1 Beispiele für Bedeutungsgleichheit**

Lassen Sie uns einige Beispiele betrachten:

- Die Aktiv-Passiv-Konvertierung: "Der Hund jagt die Katze" und "Die Katze wird vom Hund gejagt" bedeuten das Gleiche, obwohl sie sprachlich verschieden sind.
- Die Übersetzung: "Der Hund ist schwarz" hat die gleiche Bedeutung wie "The dog is black". Obwohl jedes einzelne Wort vom Ausdruck her verschieden ist, drücken beide Sätze dasselbe aus.

## **57.9 Die Lösung durch künstliche Intelligenz**

In den letzten fünf Jahren hat die KI eine Lösung für diese Herausforderung gefunden. Wie Sie sich vorstellen können, war die Computerlinguistik schon seit mindestens 50 Jahren damit beschäftigt, dieses Problem zu lösen - allerdings mit nur mäßigem Erfolg. Doch jetzt ist es möglich, und das ist einer der Gründe, warum maschinelle Übersetzung heute zur Grundausrüstung von KI-Modellen gehört.

### **57.9.1 Komplexe Übersetzungen**

Moderne KI-Systeme sind in der Lage, komplexe Texte, sogar Fachtexte, adäquat in eine andere Sprache zu übersetzen. Und zwar nicht nur Wort für Wort, wie man es früher stümperhaft versucht hat, indem man jedes einzelne Wort in eine Übertragungstabelle einfügte und froh war, wenn die grammatischen Anforderungen halbwegs erfüllt wurden. Nein, heute kann ein Satz vollständig umgebaut oder sogar in Teilsätze zerlegt werden, um das Gleiche auszudrücken - genauso wie es ein guter menschlicher Übersetzer tun würde.

## **57.10 Das Training der KI-Modelle**

Embeddings sind die Grundvoraussetzung für diesen Erfolg. Sie sind ein Teil des riesigen Trainings, das die KI-Modelle durchlaufen. Für GPT 3.5 beispielsweise dauerte das Training etwa zwei Jahre und erforderte einen extremen Computeraufwand. Durch dieses Training anhand von Embeddings und vielen Textbeispielen lernen die KI-Modelle, die Frage nach Bedeutungsgleichheit erfolgreich zu beantworten.

### **57.10.1 Die Parameter der Modelle**

Das Training ist im Grunde ein Feintuning von Milliarden von Parametern - Stellschrauben, die so justiert werden müssen, dass die KI die Anforderungen an semantische Regeln richtig umsetzt. Das Trainingsziel ist dabei ganz einfach: Die Frage nach der Bedeutung zu lösen.

### **57.10.2 Trainingsdatensätze und Übersetzlitteratur**

Für das Training gibt es hervorragende Datensätze, an denen man den Erfolg messen kann. Einer der entscheidenden Datensätze sind die Klassiker der Übersetzlitteratur. Hier haben die besten Übersetzer der Welt eine literarische Quelle in einer Sprache vorgegeben und eine höchst anspruchsvolle Übersetzung als bedeutungsgleichen Ausdruck zugeordnet. Alles, was die großen Konzerne an Übersetzlitteratur bekommen konnten, haben sie für das Training verwendet.

Das ist eines der Geheimnisse, warum nun plötzlich auch Latein gut übersetzt wird. Die KI-Entwickler haben die Klassiker der Teutner-Serie genommen, die hervorragenden Übersetzungen der Philologen, und hatten damit eine präzise Übersetzungszuordnung zwischen modernen Sprachen und den Texten von Horaz oder Cicero.

## **57.11 Weitere Trainingsdaten**

Natürlich hat man auch die gesamte philosophische Literatur digitalisiert, denn hier finden sich wertvolle sprachphilosophische Reflexionen über die Inhalte. Was sind logische Schlussformen? Das kennen Sie alles aus den Logik-Lehrbüchern. Sie können aus den KI-Programmen herauskitzeln, dass sie diese Texte Satz für Satz trainiert haben. Nicht nur Philosophie-Studenten üben in Logik 1 die Logiktexte, auch alle KI-Modelle haben das intus, weil hier die Regeln der Semantik geübt werden. Ein Modus ponendo ponens gehört zum Repertoire des Schließens für KI-Modelle genauso wie für einen Philosophie-Studenten.

Allerdings geht das manchmal auch noch deutlich daneben...# Bedeutungsgleichheit und Embeddings in der KI

In der Welt der künstlichen Intelligenz spielen Wiederholungen eine ebenso entscheidende Rolle wie beim menschlichen Lernen. So wie ein einzelner Besuch einer Logikvorlesung nicht ausreicht, um die Materie vollständig zu beherrschen, benötigen auch Maschinen mehrfache Wiederholungen, um Inhalte zu verinnerlichen. Embeddings, numerische Repräsentationen von Wörtern oder Sätzen, dienen hierbei als Grundlage für das Training von KI-Modellen zur Bedeutungszuordnung. Doch Vorsicht: Embeddings allein reichen nicht aus, um die Bedeutung sprachlicher Ausdrücke vollständig zu erfassen.

## **57.12 Kontextabhängigkeit der Bedeutung**

Die Frage, ob Ausdrücke semantisch gleich sind, lässt sich in den meisten Fällen nicht pauschal beantworten. Der Kontext, in dem Sprache verwendet wird, spielt eine entscheidende Rolle bei der Beurteilung der Bedeutung sprachlicher Vorkommnisse. Embeddings reduzieren die komplexen Bedeutungsdimensionen auf einige Tausend mathematische Dimensionen - eine starke Vereinfachung, die jedoch den aktuellen Stand der Technik widerspiegelt.

## **57.13 Funktionsweise der KI bei inhaltlichen Fragen**

Stellen Sie sich vor, Sie fragen eine KI: "Fliegt da ein Schwan über den Baum?" Die KI übersetzt diese Eingabe zunächst in eine numerische Repräsentation, die sogenannten Embeddings. Dieser Satz erhält dann eine Zahl in 5.536 Dimensionen zugeordnet - eine erstaunlich kompakte Darstellung der dahinterstehenden Komplexität. Mit dieser Zahl durchsucht die KI eine Datenbank nach bedeutungsähnlichen Aussagen, unabhängig von der Sprache oder syntaktischen Transformationen wie Aktiv-Passiv-Konstruktionen. Die Zeichenabfolge (Strings) spielt keine Rolle mehr; es geht einzig um den Inhalt.

## **57.14 Erweiterung der Embeddings auf multimediale Inhalte**

Die Revolution der KI beschränkt sich nicht nur auf Texte. Embeddings lassen sich auch auf Bilder, Videos, Audio, 3D-Objekte und sogar Hologramme anwenden. KI-Programme können somit nicht nur Texte inhaltlich verstehen, sondern auch begleitende visuelle Elemente wie Diagramme oder Daten erschließen. Diese Erweiterung eröffnet völlig neue Möglichkeiten der Informationsverarbeitung.

# **58 Attention is all you need - die zweite Revolution**

Der Artikel “Attention is all you need”, erschienen auf dem Preprint-Server arXiv der Cornell University, markiert einen weiteren Meilenstein in der KI-Revolution. Die Autoren, darunter Jakob Uskoreits Sohn, haben sich ihr ganzes Leben mit Übersetzungen beschäftigt - ein Bereich, der den Boden für die KI-Revolution bereitet hat.

## **58.1 Transformation von Sequenzen**

Der Artikel befasst sich mit der Transformation von Sequenzen, also Satzabfolgen. Sprache wird hier als eine Abfolge von Satztokenwörtern verstanden. Die Aufgabe besteht darin, nicht nur die Bedeutung dieser Ausdrücke zu identifizieren, sondern auch vorherzusagen, welches Wort als nächstes folgen könnte. Diese Funktion kennen Sie vielleicht von Rechtschreibkorrekturprogrammen oder der Wortvervollständigung auf Smartphones.

## **58.2 Die Bedeutung der Sequenztransformation**

Die Fähigkeit, die nächste Zeichenfolge vorherzusagen, mag auf den ersten Blick technisch interessant, aber nicht besonders aufregend erscheinen. Doch genau hier liegt der Schlüssel zur zweiten Revolution. Denn es geht nicht nur um sprachliche Ausdrücke, sondern um die Dimension der Bedeutung.

# 59 Rettungsversuch und KI-Demonstration

Nachdem meine Folie sich unbeabsichtigt geschlossen hat, versuche ich einen Rettungsversuch mit Hilfe der KI. Nicht, um nach einer Lösung zu fragen, sondern um Ihnen live zu demonstrieren, was der Clou an der Vorhersage der nächsten Zeichenfolge ist.

Ich gebe den Satz “Der Hund ist schwarz.” ein. Was die KI als Fortsetzung vorschlägt, ist jedes Mal anders und oft überraschend. In der Evolution des User-Interfaces ist das, was früher der Go# Interaktion mit KI-Modellen

Heute möchte ich Ihnen von einem faszinierenden Experiment berichten, das ich kürzlich mit einem KI-Modell durchgeführt habe. Stellen Sie sich vor, Sie geben dem Programm eine einfache Aussage ein, wie beispielsweise “Der Hund ist schwarz.” Was erwarten Sie, dass das Modell darauf antwortet? Genau das habe ich ausprobiert und die Ergebnisse waren höchst aufschlussreich.

## 59.1 Die Herausforderung der Feststellung

Als ich die Aussage “Der Hund ist schwarz.” in das KI-Modell eingab, war ich gespannt, welche Reaktion es zeigen würde. Zu meiner Überraschung schien das Programm zunächst etwas perplex zu sein. Es wusste offenbar nicht so recht, was es mit dieser schlichten Feststellung anfangen sollte.

Ich wiederholte die Eingabe, doch das Modell reagierte erneut mit einer Entschuldigung und gab dann eine ausführliche Beschreibung eines schwarzen Labradors aus. Es assoziierte alles, was man mit schwarzen Hunden in Verbindung bringen könnte und generierte einen regelrechten pseudoliterarischen Erguss.

Als ich die Aussage ein weiteres Mal wiederholte, erkannte das Programm immerhin, dass seine vorherige Antwort wohl nicht ganz das Richtige war. Es entschuldigte sich erneut und wiederholte die ursprüngliche Aussage in ihrer prägnanten Form.

## **59.2 Von der Frage zur Anweisung**

Dieses Experiment verdeutlicht einen wichtigen Wandel in der Nutzung von KI-Modellen. Anfangs dienten sie hauptsächlich dazu, Fragen zu beantworten - ähnlich wie bei einer Google-Suche. Die Pragmatik des Dialogführers war klar: Frage und Antwort.

Doch mittlerweile hat sich der Fokus verschoben. Statt Fragen zu stellen, geben wir den Modellen immer häufiger Anweisungen oder Instruktionen. Deshalb wurden viele Modelle neu trainiert und tragen nun Namen, die auf ihre Fähigkeit zur Ausführung von Anweisungen hindeuten.

## **59.3 Die Bedeutung der Aufmerksamkeit**

Der Attention-Mechanismus spielt bei der Generierung plausibler Textfolgen eine entscheidende Rolle. Je nachdem, ob wir eine Frage stellen, eine Instruktion geben oder etwas sagen, das offensichtlich eine bestimmte Reaktion erfordert, passt sich die Ausgabe des Modells an.

Doch was passiert, wenn wir dem Programm eine Aussage präsentieren, auf die es keine sinnvolle Antwort geben kann? Hier zeigt sich eine interessante Eigenschaft der meisten KI-Modelle: Sie sind so programmiert, dass sie immer etwas ausgeben müssen. Schweigsame Modelle, die bei einer Unsinnfrage einfach nichts sagen, gibt es nicht.

## **59.4 Die Macht der Kontextualisierung**

Der revolutionäre Aspekt der gegenwärtigen KI-Modelle liegt in ihrer Fähigkeit, sprachliche Ausdrücke zu kontextualisieren. Nehmen wir das Beispiel der Übersetzung. Wenn ich dem Programm die Anweisung gebe: “Übersetze den Text ‘Der Hund ist schwarz’”, dann versteht es die Bedeutung und gibt korrekt “The dog is black” aus.

Dieser Komplex aus Anweisung und sprachlichem Ausdruck wird vom Modell richtig verstanden und die entsprechende Ausgabe generiert. Intern reformuliert das Programm die Eingabe in eine explizite Wiedergabe des Inhalts, um sicherzustellen, dass alles eindeutig ist. # Erweiterung des Eingabekontexts zur Steuerung der Ausgabe

In den gegenwärtigen KI-Modellen wird der Kontext explizit gemacht und mit in den Eingabekontext geschrieben. Auf diese Weise wird der generierte Text gesteuert. Wenn ich beispielsweise ohne weiteren Kontext den Satz “Der Hund ist schwarz” eingebe, fängt das Programm von sich aus an, weitere Informationen zu generieren. Durch Zusatzinformationen im Kontext lässt sich die Ausgabe jedoch stark beeinflussen.

## **59.5 Das Problem der Halluzination**

Die Halluzination entsteht dadurch, dass es keinerlei Beschränkungen auf den Inhalt oder sachliche Prüfungen gibt. Die aktuellen Modelle beherrschen lediglich die Übersetzung von sprachlichem Ausdruck in ihre Bedeutung - sie haben Sprachkompetenz, aber keinerlei Sachkompetenz. Es existieren keine Mechanismen, die prüfen, ob ein generierter Satz tatsächlich sachlich korrekt ist.

Obwohl durch die KI-Programme die Dimension der Wahrheit, Rechtfertigung und Kritik eröffnet wird, lösen sie diese Fragen noch nicht ein. Sachliche Korrektheit wird nicht geprüft, Evidenzen nicht angeführt und Kritik nicht geübt. Diese Aspekte sind schlichtweg nicht Teil der Programme.

## **59.6 Konsequenzen für die Verwendung von KI-generierten Texten**

Hausarbeiten sollten niemals mit Chat-GPT geschrieben werden, da die Wahrscheinlichkeit für falsche Informationen extrem hoch ist. Die Programme sind perfekt darin, Ausgaben sinnvoll erscheinen zu lassen, aber nicht in der Lage, deren Wahrheitsgehalt zu überprüfen.

Ein Beispiel hierfür ist meine Erfahrung mit einer Abfrage zu Leonhard Eulers Publikationsverhalten im Jahr 1756. Statt zuzugeben, keine Informationen zu haben, generierte das Programm eine perfekt aussehende, aber völlig erfundene Literaturangabe. Selbst als Experte konnte ich die Fälschung zunächst nicht erkennen - so überzeugend war die Formatierung bis hin zu passenden bibliografischen Details. Ohne Fachwissen wäre der Fake nicht aufgefallen.

# **60 Erweiterung der KI-Modelle um Wissen und Validierung**

## **60.1 Notwendige Ergänzungen für sachliche Korrektheit**

Um zu garantieren, dass generierte Informationen richtig sind, müssen den KI-Modellen zusätzliche Elemente hinzugefügt werden. Sie benötigen Zugriff auf das Wissen der Welt, das ihnen momentan fehlt.

Als Wissenschaftler würde man zur Prüfung einer Aussage wie folgt vorgehen:

- Konsultation glaubwürdiger Referenzen
- Recherche in Primärquellen (z.B. Eulers Opera Omnia)
- Aufsuchen einer Bibliothek zur Verifizierung der Publikation

Dieses Vorgehen müsste in zukünftigen KI-Systemen abgebildet werden, um sachliche Korrektheit herzustellen.

## **60.2 Verhältnis von Sprache und Sachlichkeit**

Sprache und Sachlichkeit sind vergleichbar komplex. Der sprachliche Ausdruck sollte idealerweise dem sachlichen Inhalt in der Welt entsprechen - eine Korrespondenztheorie der Wahrheit. Stimmen beide überein, ist die Aussage wahr, ansonsten falsch.

Diese Korrespondenz muss den KI-Modellen methodisch beigebracht werden, um über reine Sprachkompetenz hinauszugehen. Aktuell fehlt ihnen der Zugriff auf die Realität, auf die sich die Sprache beziehen sollte.

# **61 Auswirkungen der KI-Entwicklung auf Sprache und Bedeutung**

## **61.1 Zirkularität der Bedeutung in KI-trainierten Texten**

Wenn immer mehr Texte von KI-Systemen generiert werden, die auf jahrhundertealten Daten trainiert wurden, entsteht die Gefahr einer Zirkularität der Bedeutung. Neue Bedeutungsebenen von Begriffen könnten nicht mehr hinzukommen, die Sprache käme zum Stillstand - zumindest bei Systemen, die nur reproduzieren, was sie in den Vorlagen finden.

## **61.2 Übersetzungsfähigkeiten aktueller Programme**

Die aktuellen Programme arbeiten bereits auf der Bedeutungsebene und sind in der Lage, beliebige Sätze zu übersetzen, auch wenn die übersetzte Formulierung nirgends in der Literatur vorhanden ist.

Selbst anspruchsvolle Texte wie Goethes Faust oder Werke von Thomas Mann, die nicht in jede Sprache übersetzt wurden, können von den Programmen übertragen werden. Ob die Übersetzung angemessen ist oder Fehler enthält, lässt sich diskutieren - aber die Programme werden einen Übersetzungsvorschlag liefern. # Einleitung

Einen schönen guten Morgen, meine Damen und Herren. Heute möchte ich Ihnen etwas über die faszinierenden Entwicklungen im Bereich der Künstlichen Intelligenz und insbesondere der Sprachmodelle erzählen. Lassen Sie uns gemeinsam ergründen, wie diese Systeme funktionieren und welche Herausforderungen und Möglichkeiten sich daraus ergeben.

## 62 Die Bedeutung der Sprachverwendung

Zunächst einmal stellt sich die Frage, wie die Bedeutung in der Sprache eigentlich entsteht. In der Philosophie gibt es dazu verschiedene Ansätze, aber eine zentrale Erkenntnis ist, dass die Bedeutung eng mit der tatsächlichen Verwendung der Sprache verknüpft ist. Die Sprachmodelle der KI versuchen genau das zu erfassen - sie analysieren riesige Textkorpora und lernen daraus, in welchen Kontexten bestimmte Ausdrücke typischerweise vorkommen.

Aber bedeutet das nun, dass die Entwicklung der Sprachmodelle davon abhängt, dass immer mehr Texte produziert werden? Nein, so einfach ist es nicht. Die Trainingsdaten sind in der Regel nur eine repräsentative Teilmenge aller verfügbaren Texte. Und natürlich verändert sich Sprache auch im Laufe der Zeit. Die historische Entwicklung von Bedeutungsverschiebungen ist eine große Herausforderung für die Forschung. Hier gibt es noch viel zu tun.

## 63 Die Gefahren fehlerhafter Kontexte

Ein interessantes Phänomen, das wir beobachten konnten, sind Sprachmarotten, die in den Modellen entstehen können. In einem Fall wurden die Modelle offenbar mit Texten trainiert, in denen es nicht um tatsächliche Kausalzusammenhänge ging, sondern darum, was Personen glauben, was die Ursache von etwas ist. Das führte dazu, dass die Modelle Unsinn produzierten, wenn es um kausales Schließen ging.

Dieses Beispiel zeigt, wie wichtig die sorgfältige Auswahl und Aufbereitung der Trainingsdaten ist. Fehlerhafte oder irreführende Kontexte können sich hartnäckig in den Modellen festsetzen. Manchmal hilft es schon, in den Eingabetexten explizit klarzustellen, welche Art von Antwort man erwartet. Aber in manchen Fällen muss man vielleicht auch einsehen, dass das Modell für bestimmte Aufgaben einfach nicht geeignet ist.

## **64 Erfolge und Anwendungen**

Trotz dieser Herausforderungen gibt es aber auch beeindruckende Erfolge zu vermelden. Übersetzungen waren nicht nur ein kultureller Gewinn, sondern auch ein wichtiger Motor für das Training von Bedeutungsgleichheit. Die Modelle sind inzwischen in der Lage, hochwertige Zusammenfassungen von langen Texten zu erstellen. Mit einer Kontextlänge von 200.000 Wörtern kann man ganze Bücher eingeben und sich die Kapitel in kompakter Form zusammenfassen lassen.

## 65 Die Bedeutung des Chats

Ein faszinierender Aspekt, der oft übersehen wird, ist die Rolle des Chats in der Interaktion mit Sprachmodellen wie ChatGPT. Die Entwickler wissen natürlich, dass man Begriffe nicht einfach durch notwendige und hinreichende Definitionen eingeben kann. Stattdessen nutzen sie Wittgensteinsche Gebrauchsdefinitionen. Und genau hier kommt der Chat ins Spiel.

In einem Chat findet oft eine Art semantische Korrektur statt. Wir fragen "Was meinst du damit?" oder "Meinst du dies oder jenes?". Dadurch klären wir die Bedeutung dessen, was der andere gesagt hat. Und genau das passiert auch in der Interaktion mit ChatGPT. Wenn wir eine Frage stellen und das Modell antwortet, dann können wir durch Rückfragen und Klarstellungen den Kontext präzisieren.

Das bedeutet, dass wir als Nutzer aktiv zur Intelligenz des Systems beitragen. Indem wir chatten, helfen wir dem Modell, die Bedeutung zu erschließen und bessere Antworten zu geben. Das ist ein genialer Ansatz, der bewusst so entworfen wurde und bis heute genutzt wird.  
# Begrüßung und organisatorische Hinweise

Herzlich willkommen zur dritten Vorlesung unserer Reihe "Philosophie der AI". Bevor wir in die inhaltliche Diskussion einsteigen, gestatten Sie mir noch ein paar organisatorische Anmerkungen, insbesondere zur Testatvergabe für ÜWP-Studenten.

Das Agnes-Zulassungssystem, das eigentlich gar nicht Teil des Matrikulationssystems der Humboldt-Universität ist, spielt leider immer wieder verrückt und lehnt willkürlich, ohne Rücksprache mit mir oder der Fakultät, Bewerbungen oder Meldungen zur Vorlesung ab. Einige von Ihnen haben solche Ablehnungen erhalten, aber lassen Sie sich davon nicht beirren. Das sind lediglich unmotivierte Systemreflexe, die Sie getrost ignorieren können. Die Zulassung zu dieser Vorlesung erfolgt durch die Fakultät und mich persönlich. Solange wir Platz in diesem Vorlesungssaal haben, ist jeder immatrikulierte Student, auch ÜWP-Studenten, herzlich willkommen.

Wichtig ist nur, dass wir sicherstellen, dass Ihre Studienleistung letztendlich auch in das Prüfungssystem Ihrer jeweiligen Fakultät eingetragen wird. Hierbei ist Ihr Hauptfach federführend für die Administration Ihrer Leistungsnachweise verantwortlich. Sollten Sie also noch Bedenken oder Nachfragen haben, wenden Sie sich bitte direkt an das Prüfungsbüro Ihres Hauptfaches. Diese werden in der Regel bei ÜWP-Studenten die Anfrage an die jeweiligen Nebenfächer delegieren - so zumindest die gängige Praxis an der Philosophischen Fakultät.

Zuständig für die Administration ist Frau Krause vom Sekretariat der Philosophie. Sie trägt die Leistungen ein, obwohl es eigentlich bereits einen Beschluss sowohl der Fakultät als auch

anderer Gremien gibt, dass wir keine individuellen Einträge für ÜWP-Studenten mehr benötigen. Irgendwie geht das alles ein bisschen durcheinander. Im letzten Semester konnten Sie sich selbst eintragen und melden, dieses Semester haben Sie Ablehnungsscheine und völlig unmotivierte Ablehnungsbescheide erhalten, die aber, wie gesagt, irrelevant sind. Es ist leider etwas konfus, aber die Kernbotschaft ist: Sie sind alle offiziell zugelassen! Wir müssen nur sicherstellen, dass Ihre Leistungstestdate in Ihren entsprechenden Prüfungsbögen dokumentiert werden. Hierzu möchte ich Sie bitten, noch kurz Rücksprache zu halten. Der direkte Weg, was unsere Fakultät betrifft, wäre eine E-Mail an Frau Krause vom Sekretariat der Philosophie. Diesen Weg gehen wir. Falls im Laufe des Semesters noch weitere Rückmeldungen kommen, was Sie alles tun müssen, melden Sie sich einfach bei mir. Wir bekommen das schon geregelt, auch wenn es heute etwas kompliziert und aufwendig erscheint.

## **66 Mögliche Projektarbeiten**

In etwa einer halben Stunde werde ich noch etwas über die möglichen Projektarbeiten sagen, die Sie im Laufe dieser Vorlesung als zusätzliche Leistungspunkte absolvieren können. Diese Projektarbeiten sind integriert in ein spannendes Gesamtforschungsvorhaben, das derzeit in Kooperation mit zwei wissenschaftlichen Akademien und der Stiftung Deutscher Klassik in Weimar organisiert wird.

Wir überlegen noch - und das hängt auch ein bisschen von Ihnen und Ihrem Engagement ab - ob die Ergebnisse, die aus den von uns konzipierten Übungen hervorgehen, zum Ende des Semesters öffentlichkeitswirksam dokumentiert und mindestens auf einer Webseite präsentiert werden. Was genau das sein wird, hoffe ich Ihnen im Laufe dieser Vorlesung unterbreiten zu können. Falls Sie dazu Nachfragen oder Klärungsbedarf haben, melden Sie sich am besten gleich während der Vorlesung, sodass wir das direkt ausdiskutieren können.

# 67 Generative Modelle der AI

Kommen wir nun zum eigentlichen Thema unserer heutigen Vorlesung. In den letzten beiden Stunden haben wir bereits etwas vertieft die verschiedenen Modelle der AI oder KI diskutiert. Einer der Begriffe, der sich in diesem Zusammenhang zunehmend etabliert, ist der der generativen Modelle der AI.

Generativ sind diese Modelle schlichtweg deshalb, weil sie in der Lage sind, Texte zu erzeugen. Sie generieren also etwas, und zwar abhängig von einem Input, den sie erhalten. Heute werden solche Modelle verstärkt auch zur Generierung von Bildern, Videos oder Audiodaten genutzt. Das generelle Schema ist denkbar einfach: Irgendetwas wird eingegeben, sei es über Tastatur, Sprache oder das Einlesen von Datenbeständen. Diese Eingabe wird dann von den Modellen verarbeitet und es erfolgt eine entsprechende Ausgabe. Das ist alles, was hinter dem Begriff "generative Modelle der AI" steckt.

Ein anderer Begriff, der in der Diskussion und Literatur häufig verwendet wird und sich zunächst sehr kompliziert anhört, im Grunde aber das Gleiche meint, ist die Abkürzung LLM, die für "Large Language Models" steht. Wie der Name schon sagt, handelt es sich hierbei um große Sprachmodelle, die vom Umfang her beträchtlich sind. Die Modelle, die generative AI ausmachen, sind im Kern Sprachmodelle, auch wenn sie sogar in der Lage sind, Bilder zu verarbeiten und zu interpretieren.

Diese Konzeption ist sehr umfassend und wir werden hoffentlich im Laufe der Vorlesung noch besser verstehen, warum die Ebene des Sprachverarbeitens und Sprachverständnisses so zentral für alle Modelle der künstlichen Intelligenz ist, mit denen wir es hier zu tun haben.

## 67.1 Vielzahl verschiedener Modelle

Es gibt derzeit etwa 100 verschiedene Modelle dieser Art, wobei die Zahl stetig wächst. Manche sind nur über Lizzenzen und Zugriffsbarrieren nutzbar, die Mehrzahl ist mittlerweile aber Open Access verfügbar. Es ist bemerkenswert, wie die Zahl der angebotenen Modelle mit jeweils unterschiedlichen Kompetenzen förmlich explodiert.

Dabei ist es nicht so, dass es immer nur das Gleiche ist, nur unter einer anderen Überschrift von entsprechenden Trägern oder Entwicklern. Vielmehr haben die verschiedenen Modelle durchaus unterschiedliche Fähigkeiten, auf die wir gleich noch näher eingehen werden.

## 67.2 Funktionsweise der aktuellen Modelle

Entscheidend ist, dass die derzeitige Generation der Modelle so arbeitet, dass sie auf der einen Seite ein Large Language Model, also ein Sprachmodell haben, das im Wesentlichen so zu verstehen ist, dass es an riesigen Corpora von Übersetzungstexten trainiert wurde, um zwei Kernaspekte der KI-Revolution zu beherrschen:

1. Semantische Ähnlichkeit feststellen: Das heißt, die Modelle sind in der Lage, Bedeutungsgleichheiten oder Bedeutungsähnlichkeiten zu identifizieren. Es geht also nicht um ein simples semantisches Matchen wie bei einer Google-Suche nach Stichworten, sondern um das Erkennen inhaltlicher Bedeutungsähnlichkeiten.
2. Transformation anwenden: Die Transformation sorgt dafür, dass die erkannten Ähnlichkeiten genutzt werden, um bei einem gegebenen Input einen passenden Output zu generieren.

Es ist die Kombination dieser beiden Aspekte, die den entscheidenden Unterschied ausmacht. Würde man nur nach Modellen suchen, die aufgrund eines Textinputs - ähnlich einer Abfrage oder Suche, wie wir es von Google kennen - ohne Berücksichtigung semantischer Ähnlichkeiten arbeiten, würde man lediglich entsprechende Strings, also Textfolgen oder Zeichenketten finden, mehr nicht.

Die Kombination von Transformation und Ähnlichkeitserkennung ermöglicht es den Modellen, etwas extrem Weitreichendes zu tun, nämlich aufgrund der Bedeutungsähnlichkeiten der Textinhalte auch die Transformation als Regeln in verallgemeinerter Art und Weise anzuwenden.

Vergleichbar ist das mit einer allgemeinen Regel wie "Bei schönem Wetter über 26 Grad können wir draußen zu Mittag essen". Das ist eine verallgemeinerte Regel, die etwas über unser Verhalten aussagt, unabhängig von einem spezifischen Tag, Monat oder Ort. Und genau solche verallgemeinerten Regeln sind diesen Modellen bekannt, wenn sie in der Lage sind, die beiden genannten Aspekte zu kombinieren: Die Semantik, also die Bedeutung der Ausdrücke zu verallgemeinern und die Transformation dieser Regeln zu generieren.

Deswegen ist die Kombination dieser beiden Fähigkeiten so extrem weitreichend und macht den entscheidenden Unterschied zu früheren Ansätzen aus.<sup>#</sup> Erweiterte Möglichkeiten durch Charakterdefinition bei KI-Modellen

Meine Damen und Herren, lassen Sie uns heute einen faszinierenden Blick in die Welt der künstlichen Intelligenz werfen. Wir haben uns bisher intensiv mit den revolutionären Komponenten der generativen KI beschäftigt - der semantischen Gemeinierung und der Transformation auf Regeln. Doch jetzt kommt ein drittes Element hinzu, das die Philosophie ins Spiel bringt und uns in den kommenden Vorlesungen noch eingehender beschäftigen wird: die Charakterdefinition.

## **67.3 Kommunikation mit KI-Modellen wie mit einer menschlichen Person**

Stellen Sie sich vor, Sie interagieren mit einem KI-Modell, das Ihnen textlich auf eine Art und Weise antwortet, als würden Sie mit einer menschlichen Person über eine Schnittstelle kommunizieren. Und zwar nicht nur auf standardisierte Art und Weise, sondern mit einer Flexibilität und Anpassungsfähigkeit, die uns in Staunen versetzt. Durch die Beherrschung der semantischen Gemeinierung und der Transformation auf Regeln können wir die Art und Weise der Regelerstellung dieser Modelle modifizieren. Wir können sie charakterlich formen, sodass ihre Antworten uns wie Charaktereigenschaften einer Person erscheinen.

## **67.4 Stilistische Anpassungsmöglichkeiten**

Die Möglichkeiten dieser charakterlichen Formung sind schier grenzenlos. Wir können die Sprache, den Schreibstil und sogar den literarischen Stil der Antworten beeinflussen. Stellen Sie sich vor, Sie bitten das Modell, Ihnen im Stil von Ernest Hemingway zu antworten. Obwohl es sich natürlich um ein Imitat handelt, da Hemingway nicht mehr lebt, können die Modelle eine beeindruckende Stilähnlichkeit herstellen. Sie können die Datenausgabe knapper formulieren, schematisieren oder in bestimmten Datenformaten ausgeben lassen. Sogar griechische Hexameter oder der Stil eines Homer sind möglich. Mit dieser Technik wären wir der alten Forderung, Universitätsdissertationen auf Latein abzugeben, schon ziemlich nahe - und die Modelle würden vermutlich weniger Lateinfehler produzieren als die Doktoranden vor 100 Jahren.

## **67.5 Konfiguration formaler inhaltlicher Regeln**

Doch die wahre Faszination liegt in der Konfiguration formaler inhaltlicher Regeln des Nachdenkens, des Resonierens und des Formulierens von Arbeitsverfahren und Denkprozessen. Welche Aspekte müssen berücksichtigt werden, um die Instruktionen, die wir den Modellen geben, korrekt zu erfüllen? Dieser Bereich ist in der derzeitigen Entwicklung der Technologie noch sehr unterbelichtet, obwohl alle Entwickler wissen, dass solche Aspekte befolgt werden müssen. Der Forschungsbereich des Reasonings ist einer der am intensivsten durchgeführten und am schwersten finanzierten Bereiche. Wir werden gleich kennenlernen, warum das so relevant ist.

## **67.6 Charakterdefinition als zusätzliche Dimension**

All diese Aspekte - die stilistischen Anpassungen, die formalen inhaltlichen Regeln und die Metaregeln - fasste ich unter dem Begriff der Charakterdefinition oder Typ-Einstellung des jeweiligen Modells zusammen. Diese Dimension kommt zusätzlich zu den bereits vorhandenen Sprachkompetenzen hinzu. Man kann sich das so vorstellen, als hätte man es mit einem Menschen zu tun, der hervorragend Sprachen gelernt hat, aber sonst nichts. Genau so verhält es sich derzeit im Wesentlichen mit den KI-Modellen.

## **67.7 Herausforderungen und Zukunftsaufgaben**

Es gibt noch viele Herausforderungen zu meistern. Beschränkungen moralischer Art, beispielsweise das Artikulieren von Unverschämtheiten, kennen die Modelle noch nicht. Einige sanktionieren solche Äußerungen und geben dann schlichtweg nichts aus, was aber oft als ärgerliche Zensur empfunden wird. Diese ganzen Dimensionen stehen noch am Anfang und gehören zur umfassenden Konfiguration eines künstlichen Charakters.

Auch Metaregeln spielen eine große Rolle, insbesondere in Bereichen wie dem medizinischen kausalen Schließen. Wie ist eine medizinische Diagnostik durchzuführen? Welche kausalen Vorstellungen über Krankheiten, Krankheitsverläufe und Diagnostik gehören dazu? Die Modelle werden zwar über riesige Mengen an Publikationen trainiert, aber allgemeine Metaregeln zum korrekten Schließen und zu wissenschaftlichen Verfahren abzuleiten, ist noch nicht so einfach. Dieses Training ist eine der enorm wichtigen Zukunftsaufgaben, die es zu lösen gilt.

## **67.8 Historisches Schließen als Anwendungsbeispiel**

Ein weiterer interessanter Bereich, den ich Ihnen anhand von Fallstudien vorstellen möchte, ist das historische Schließen. Wenn man historische Aussagen über Biografien bekannter Persönlichkeiten trifft, darüber, was und wann sie an Ereignissen erlebt und darüber berichtet haben oder wer wen geprägt hat - das sind typische Aufgabenbereiche in den historischen Wissenschaften. Auch hier gibt es Regeln, wie damit umzugehen ist. Diese Regeln müssen den Programmen noch beigebracht werden, da sie bisher nur durch Beispiele gelernt haben, einen kleinen Bereich anzuwenden. Ich bin jedoch sicher, dass diese Probleme innerhalb der nächsten zwei Jahre gelöst sein werden und die entsprechenden Leistungen dann verfügbar sind.

## **67.9 Bedeutung des Kontexts**

Neben den Regeln spielt auch der sogenannte Kontext eine entscheidende Rolle für den Input der Transformation von generativen Modellen. Unter Kontext versteht man in dieser Terminologie alle Informationen, sprich Sätze und Texte, die das Programm zusätzlich zu einer bestimmten Instruktion als Eingabe erhält, um einen entsprechenden Output zu generieren. Dieser Kontext ist extrem wichtig, um eine spezifische Aufgabe inhaltlich korrekt zu verstehen. Die Größe des Kontexts ist derzeit eine der interessantesten technischen Herausforderungen. Es gilt, den Kontext maximal groß zu gestalten, ohne dabei die Größe des Modells exponentiell wachsen zu lassen. Andernfalls würden die Anforderungen an Hardware, Software und Strom zu groß werden und die Bearbeitungsdauer durch die Modelle zu lang, was die Praktikabilität einschränken würde.

Lassen Sie uns gemeinsam die Entwicklung dieser faszinierenden Technologie verfolgen und erkunden, welche Möglichkeiten sich durch die Charakterdefinition bei KI-Modellen eröffnen. Es liegt noch ein spannender Weg vor uns, aber ich bin zuversichtlich, dass wir in den kommenden Jahren bedeutende Fortschritte erleben werden. # Einleitung

Sehr geehrte Damen und Herren,

heute möchte ich Ihnen einen tieferen Einblick in die faszinierende Welt der künstlichen Intelligenz (KI) geben und insbesondere darauf eingehen, welche Rolle der Kontext und die Sachkompetenz bei der Entwicklung leistungsfähiger KI-Modelle spielen. Wir werden sehen, dass die Integration zusätzlicher Informationen zwar einerseits das Potenzial hat, die Fähigkeiten der KI enorm zu steigern, andererseits aber auch mit großen Herausforderungen verbunden ist.

## **68 Kontext und Sachkompetenz als Schlüsselfaktoren**

Je mehr Kontext ein KI-Modell berücksichtigen kann, desto besser wird es in der Lage sein, komplexe Aufgaben zu bewältigen und fundierte Entscheidungen zu treffen. Doch der Preis dafür ist hoch: Die Verarbeitung riesiger Datenmengen erfordert gewaltige Ressourcen an Zeit, Geld und Rechenleistung. Es ist ein schmaler Grat zwischen dem Streben nach immer leistungsfähigeren Systemen und der Notwendigkeit, die Kosten im Rahmen zu halten.

Die Sachkompetenz ist ein weiterer entscheidender Faktor, der bisher jedoch noch nicht ausreichend in die KI-Modelle integriert werden konnte. RAG (Ressource) ist ein Ansatz, bei dem zusätzliche Informationen als Input für den Kontext bereitgestellt werden, um die Sachkompetenz zu verbessern. Trotz intensiver Forschung in den letzten Monaten sind die bisherigen Lösungen jedoch oberflächlich und nicht zufriedenstellend.

## **69 Die Grenzen der KI**

Es ist wichtig zu verstehen, dass die beeindruckenden Ergebnisse, die KI-Modelle liefern, oft täuschend sein können. Sie klingen überzeugend und plausibel, sind aber in Wirklichkeit keiner echten Sachprüfung unterzogen worden. Die Wahrscheinlichkeit von Fehlern ist hoch, und Nutzer sollten sich dessen bewusst sein.

## **70 AGI - Der Traum von der Superkompetenz**

In der Debatte um die Zukunft der KI taucht immer wieder der Begriff AGI (Artificial General Intelligence) auf. Manche sehen darin die Krönung der kognitiven Kompetenz, die alle menschlichen Fähigkeiten übersteigt. Doch ich halte diese Sichtweise für fragwürdig. Es gibt bereits heute Bereiche, in denen Maschinen dem Menschen weit überlegen sind, etwa bei der Lösung komplexer mathematischer Probleme. Doch das bedeutet nicht, dass sie in allen Bereichen die Oberhand gewinnen werden.

# 71 Die Zukunft der KI

Wie wird sich die KI in Zukunft entwickeln? Das ist eine Frage, die derzeit niemand mit Sicherheit beantworten kann. Die Dynamik in diesem Bereich ist so groß, dass seriöse Prognosen über einen längeren Zeitraum kaum möglich sind. Statt auf eine allumfassende Superkompetenz hinzuarbeiten, gehe ich davon aus, dass sich die KI in Richtung Spezialisierung entwickeln wird. Wir werden Modelle sehen, die auf bestimmte Aufgaben wie Sprachverarbeitung, Diagnostik oder mathematische Berechnungen zugeschnitten sind.

## **72 Die Bedeutung der Geisteswissenschaften**

Ein Bereich, der für die Geisteswissenschaften von besonderer Bedeutung ist, ist die Interpretation von Texten. Hier geht es darum, den Inhalt kritisch zu hinterfragen, zu verstehen und zu interpretieren. Ich bin überzeugt, dass auch Computer in Zukunft in der Lage sein werden, hermeneutische Verfahren anzuwenden und so zu einem tieferen Textverständnis zu gelangen. Doch dafür müssen die Modelle nicht nur auf der Basis von Trainingsdaten lernen, sondern auch eigene Interpretationsleistungen erbringen können.

## 73 Schlussgedanken

Die Entwicklung der KI schreitet rasant voran, und es ist schwer vorherzusagen, wohin die Reise gehen wird. Fest steht jedoch, dass die Integration von Kontext und Sachkompetenz eine der größten Herausforderungen bleibt. Nur wenn es gelingt, diese Faktoren sinnvoll in die Modelle einzubinden, werden wir in der Lage sein, das volle Potenzial der KI auszuschöpfen und gleichzeitig die Risiken zu minimieren.<sup>#</sup> Definition der Künstlichen Intelligenz

In der Vergangenheit wurde Künstliche Intelligenz oft als ein System definiert, das auf verschiedenen Ebenen, sei es Physik, Chemie oder andere Bereiche, quasi angelernt wurde und dadurch mehr Fähigkeiten erlangte. Diese Definition ist mir bekannt, aber ich möchte an dieser Stelle nicht zu sehr ins Detail gehen. Vielleicht können wir später nochmal darauf zurückkommen, wenn wir die einzelnen erforderlichen Kompetenzbereiche näher beleuchtet haben.

## 74 Bewertung der Leistungsfähigkeit von KI-Modellen

Derzeit befinden sich Hunderte von KI-Modellen in einem riesigen Wettbewerb. Wie in Amerika üblich, werden die Leistungen der Modelle in Tabellen erfasst, ähnlich wie bei Sportwettbewerben oder der Elo-Bewertung im Schach. Die Modelle müssen standardisierte Prüfungsfragen beantworten, die in verschiedenen Fachdisziplinen gestellt werden - vergleichbar mit den jährlichen Abiturfragen in Deutschland. Sogar Aufgaben der mathematischen Olympiade für Nachwuchsmathematiker, die äußerst anspruchsvoll sind, werden den Programmen vorgelegt.

Mittlerweile schneiden die trainierten KI-Modelle bei den meisten juristischen Standardaufgaben besser ab als menschliche Kandidaten im Bachelorstudium. Der durchschnittliche amerikanische Jurastudent erzielt schlechtere Ergebnisse als die entsprechenden KI-Systeme. Es gibt eine Vielzahl solcher Tests mit umfangreichen Fragesammlungen aus etwa 40 bis 50 verschiedenen Bereichen. Jedes Modell, das an diesem Wettbewerb teilnehmen möchte, muss einen automatischen Testparcours durchlaufen. Dabei wird ermittelt, wie viele Fragen in einem bestimmten Kompetenzbereich korrekt beantwortet werden. Die erreichten Prozentzahlen werden veröffentlicht und jedes ambitionierte Modell muss angeben, welchen Prozentsatz es bei einem bestimmten Fragesatz erreicht hat.

Allerdings gibt es bei diesen Tests ein Problem: Die Modelle können immer besser auf genau diese Fragen trainiert werden. Das Verfahren ist derzeit noch ziemlich unsauber, weshalb solchen Qualitätsmaßstäben nicht wirklich zu trauen ist. Interessanter sind daher ELO-Wettbewerbe, bei denen ein Modell einem anderen eine Frage stellen darf. Wenn das andere Modell die Frage löst, darf es wiederum eine Frage stellen, muss aber auch selbst die Lösung finden. Die Lösung darf nicht bereits als trainiertes Ergebnis vorliegen. Wie bei einem Schachturnier wird so ermittelt, welches Modell in welchem Kompetenzbereich häufiger gewinnt. Nach den Regeln des Schachspiels werden den Modellen dann ELO-Qualifikationen zugewiesen, die sich von Woche zu Woche ändern können.

Natürlich hängt auch dieses Ranking stark davon ab, welche Aufgaben gestellt werden. Werden bestimmte Aufgabenbereiche nicht abgefragt, werden die Modelle in diesem Sektor auch keine entsprechenden Kompetenzen entwickeln. Dies wirft interessante Fragen hinsichtlich der Bewertung der allgemeinen Intelligenz (AGI) dieser Modelle auf. Die Situation ist sehr dynamisch und die Definition von AGI hängt stark vom theoretischen Standpunkt ab, aus dem man die Leistung der Modelle beurteilt.

## 75 Kontext und Kontextgröße

Der Kontext, von dem hier die Rede ist, bezieht sich ganz einfach auf die Anzahl der Token - also Wörter und Satzzeichen - die ein Modell berücksichtigen kann, um den Sinn einer Anfrage zu verstehen. Vor einem halben Jahr lag dieser Wert bei etwa 1.000 Token, was ungefähr drei Seiten entspricht. Alles darüber hinaus wurde nicht berücksichtigt.

Wenn man beispielsweise Informationen aus Enzyklopädie-Einträgen benötigte, die oft einen Umfang von 20 Seiten haben, war es unmöglich, diese vollständig in den Kontext der Modelle einzubinden. Irgendwo musste zwangsläufig abgeschnitten werden, egal mit welchen Verfahren man Informationen ausließ - es fehlte immer etwas.

In den letzten sechs Monaten ging es daher darum, den Kontext zu vergrößern. Die Standardmodelle, die ich in dieser Vorlesung zur Veranschaulichung nutze - die Modelle der Klasse "Cloth" (geschrieben wie "Cloud" auf Englisch oder Französisch) - haben mittlerweile einen Kontext von 200.000 Wörtern. Das ist schon eine beachtliche Menge, in der viele Informationen untergebracht werden können.

Allerdings gibt es auch hier wieder Vortäuscher, die einen großen Kontext suggerieren, der faktisch aber nicht genutzt wird. Man muss immer sehr kritisch hinterfragen, ob die angegebene Kontextgröße, z.B. 200.000 Wörter, bei der Suche nach einer Antwort tatsächlich gleichmäßig berücksichtigt wird.

## 76 Der Nadeltest

Ein sehr praktischer und aussagekräftiger Test dafür ist der sogenannte Nadeltest. Die Idee dahinter ist folgende: In einem beliebigen Text, z.B. Goethes gesammelten Werken, fügt der Nutzer an einer Stelle eine selbst gewählte Formulierung ein - etwas, das Goethe so nie geschrieben hätte, wie "Trump ist blöd". Diese Formulierung wird irgendwo in "Faust 3" eingeschleust.

Die Aufgabe für das Modell besteht darin, genau diese Feststellung zu finden - allerdings nicht wortgleich, sondern inhaltlich. Es handelt sich sozusagen um die sprichwörtliche Nadel im Heuhaufen. Man weiß nur, dass Goethe irgendwo in seinen gesammelten Werken eine Äußerung zu Trump getätigt hat. Danach soll gesucht werden, wobei der Name "Trump" nicht unbedingt erwähnt wird. Es könnte auch heißen: "Der Präsident, der 2018 im Amt war in den Vereinigten Staaten, ist blöd."

Um diese Nadel im Heuhaufen zu finden, reicht es nicht aus, einen großen Textbestand zu beherrschen. Man muss nach etwas suchen, dessen Wortlaut man nicht kennt, aber dessen Bedeutung man erfassen möchte. Hier wird genau die Art von Intelligenz gefordert, von der wir sprechen.

Solche Tests werden durchgeführt, um zu sehen, ob die verwendeten Modelle tatsächlich die Kontextgröße haben, die erforderlich ist, um einen gesamten Textbestand zu durchsuchen und zu bearbeiten. Es wäre zum Beispiel nicht erlaubt, den Gesamttext in praktikable Teile zu unterteilen und nur darin zu suchen. Die Suche muss in der Gesamtheit erfolgen.

Goethes gesammelte Werke umfassen definitiv mehr als 200.000 Wörter. Das wäre eine Aufgabe, die das Leistungsvermögen der meisten, wenn nicht aller mir bekannten Modelle übersteigt. Auf diese Weise lässt sich der Kontext testen und eine Anforderung an die Lösungskompetenz der Modelle formulieren. Wird diese Hürde von einem Modell überwunden, kann man sagen, dass es diese Fähigkeit beherrscht.

Meiner Meinung nach ist ein solcher Katalog spezifischer, lösbarer Aufgaben eine wesentlich bessere Beurteilungsgrundlage für die Leistungsfähigkeit von KI-Systemen als generelle Kriterien wie AGI. Neben dem Charakter eines Modells, also seinen Einstellungen und seinem Kontext, spielen auch die eigentlichen Instruktionen eine wichtige Rolle. Wie ich in der letzten Stunde erläutert habe, handelt es sich dabei um sprachliche Ausdrücke, die als Auslöser dienen...# Die Bedeutung von Instruktionen für generative KI-Modelle

In der Welt der generativen KI-Modelle spielen Instruktionen eine zentrale Rolle. Sie sind das Mittel, mit dem wir diesen Modellen Aufgaben übertragen und sie anweisen, bestimmte Lösungen zu generieren. Doch was genau macht eine gute Instruktion aus? Und wie unterscheidet sie sich von einer einfachen Feststellung oder Frage?

## 76.1 Die Eigenschaften einer effektiven Instruktion

Eine effektive Instruktion zeichnet sich durch mehrere Schlüsseleigenschaften aus. Sie sollte:

- Suggestiv sein und nahelegen, was getan werden soll
- Auffordernd sein und klar kommunizieren, welche Aktion erwartet wird
- Spezifisch und detailliert genug sein, um eine adäquate Antwort zu ermöglichen

Ein Beispiel wie “Der Hund ist schwarz” erfüllt diese Kriterien nicht. Es ist eine vage Feststellung, die keine klare Aufforderung enthält. Die meisten generativen KI-Modelle sind darauf trainiert, immer eine Antwort zu generieren, auch wenn die Eingabe keine wirkliche Instruktion ist. Wie sie auf solch eine nicht-kommunikative Anfrage reagieren, unterscheidet sich von Modell zu Modell.

## 76.2 Von der Query zur Instruktion

Vor etwa einem Jahr waren Queries, ähnlich wie Google-Suchanfragen, noch das vorherrschende Paradigma in der Interaktion mit KI-Modellen. Doch inzwischen hat sich der Fokus auf Instruktionen verlagert - ein allgemeineres Konzept, das verschiedenste Aufgaben umfasst.

Aus philosophischer Sicht sind Instruktionen im Grunde Handlungsanweisungen. Sie richten sich an die KI-Modelle und weisen sie an, basierend auf diesen Anweisungen Lösungen zu generieren. Die Fähigkeit, Instruktionen zu verstehen und auszuführen, ist derzeit das wichtigste Maß für das Problemlösungsvermögen generativer KI-Modelle.

## 76.3 Die technischen Grundlagen

Die Ausführung von Instruktionen durch KI-Modelle basiert im Wesentlichen auf zwei Kernprinzipien:

1. Semantische Ähnlichkeit: Das Modell muss in der Lage sein, die Bedeutung der Instruktion zu erfassen und mit seinem Wissen in Verbindung zu bringen.
2. Regelhafte Textgenerierung: Basierend auf diesem Verständnis muss das Modell dann einen kohärenten, den Regeln der menschlichen Sprache folgenden Text generieren.

Es ist eine beeindruckende Leistung der modernen KI, dass sie in der Lage ist, diese komplexen Aufgaben zu bewältigen. Doch wie wir gleich sehen werden, spielt auch der Kontext eine entscheidende Rolle.

# 77 Die Bedeutung des Kontexts

Um die Bedeutung des Kontexts zu veranschaulichen, möchte ich ein praktisches Beispiel durchspielen. Stellen wir uns vor, wir stellen einem KI-Modell die Frage: "Wer war Johann Wolfgang von Goethe?"

## 77.1 Eine typische Google-Frage

Diese Frage ist ein typisches Beispiel für eine Google-Suche. Wenn wir sie in eine Suchmaschine eingeben, erwarten wir entweder direkte Links zu Webseiten, die die Antwort enthalten, oder eine von der Suchmaschine selbst aufbereitete Zusammenfassung der relevanten Informationen.

Doch was passiert, wenn wir diese Frage einem generativen KI-Modell stellen? In diesem Fall verwende ich das Modell "Claude" von Anthropic. Wenn ich die Frage eingebe, erscheint sie oben rechts in blau. Die Antwort des Modells wird darunter generiert, erkennbar am braunen Strich, der für das Claude-Modell charakteristisch ist.

## 77.2 Trainingsgrundlage und Sachkompetenz

Die Antwort, die das Modell generiert, wirkt sachlich, detailliert und informativ. Doch woher stammt diese scheinbare Sachkompetenz? Um das zu verstehen, müssen wir einen Blick auf die Trainingsdaten dieser Modelle werfen.

Grundsätzlich kann man davon ausgehen, dass alle diese Modelle auf der gesamten Wikipedia trainiert wurden. Das heißt, alle Informationen, die in der Wikipedia enthalten sind, wurden in irgendeiner Form vom Modell verarbeitet. Hinzu kommen Millionen wissenschaftlicher Publikationen von Preprint-Servieren, hauptsächlich aus den Bereichen Physik, Informatik, Mathematik, Biologie und Medizin. Ein Manko ist allerdings, dass geisteswissenschaftliche Werke in diesen Trainingsdaten oft unterrepräsentiert sind.

Zudem wurden die Modelle an Übersetzungskorpora trainiert, einschließlich deutsch-englischer Werke. Das bedeutet, dass auch die gesammelten Werke vieler großer Autoren, zu denen Übersetzungen existieren, in die Trainingsmenge eingeflossen sind.

### **77.3 Grenzen der Sachkompetenz**

Doch obwohl in diesen Trainingsdaten viel Sachinformation steckt, bedeutet das nicht, dass diese Information auch bewertet oder geprüft wird. Die Modelle haben kein System, um die Korrektheit historischer Fakten systematisch zu verifizieren.

Die scheinbar hohe Qualität der generierten Informationen stammt oft einfach daher, dass die Modelle große Mengen an Text über ein bestimmtes Thema verarbeitet haben. Da Wikipedia eine relativ verlässliche Quelle ist, führt das häufig zu korrekten Antworten. Aber wenn die Trainingsdaten Fehler enthalten, haben die Modelle keine Möglichkeit, diese zu erkennen.

# **78 Die Herausforderung der Aktualität**

Ein weiteres Problem ist die Aktualität der Daten. Die Trainingsdaten der Modelle sind statisch, das heißt, sie enden zu einem bestimmten Datum. Wenn man Fragen über aktuelle Ereignisse stellt, wird das Modell oft keine Antwort generieren können, weil diese Informationen nicht in den Trainingsdaten enthalten sind.

Die Modelle werden zwar regelmäßig mit neuen Daten aktualisiert, aber es ist oft schwierig nachzuvollziehen, wie umfassend und aktuell diese Updates tatsächlich sind. Selbst wenn ein Modell angibt, Daten bis zu einem bestimmten Monat verarbeitet zu haben, bedeutet das nicht unbedingt, dass diese Informationen auch vollständig und ausgewogen sind.

## **78.1 Die Notwendigkeit kritischer Prüfung**

Was den generierten Antworten fehlt, ist eine kritische epistemische Prüfung. Um wirklich verlässlich zu sein, müssten die Informationen nach strengen historischen Kriterien auf ihre Richtigkeit überprüft werden. Das ist eine Herausforderung, an der derzeit viele Forscher arbeiten.

Die scheinbare Kompetenz der KI-Modelle darf uns nicht darüber hinwegtäuschen, dass sie letztlich nur statistische Muster in riesigen Textmengen erkennen. Sie haben kein echtes Verständnis für die Inhalte und können die Qualität der Informationen nicht selbstständig beurteilen.

Es liegt an uns Menschen, die Antworten der KI kritisch zu hinterfragen und ihre Aussagen sorgfältig zu prüfen. Nur so können wir sicherstellen, dass die generative KI ein nützliches Werkzeug bleibt und nicht zur Quelle von Fehlinformationen wird. Es ist eine Aufgabe, die Wachsamkeit und Engagement von uns allen erfordert.<sup>#</sup> Die Herausforderungen der Nutzung von Internetquellen für KI-Systeme

Meine sehr geehrten Damen und Herren, lassen Sie uns heute über ein Thema diskutieren, das in der Welt der künstlichen Intelligenz von größter Bedeutung ist: Die Herausforderungen, mit denen KI-Systeme konfrontiert sind, wenn sie Informationen aus dem Internet nutzen.

## **78.2 Die Problematik widersprüchlicher Informationen**

In der Welt des Internets ist es nicht ungewöhnlich, auf widersprüchliche Informationen zu stoßen. Ob eine Quelle nun aus dem Dezember 2023 oder aus dem Januar 1905 stammt, spielt dabei oft keine entscheidende Rolle. Doch genau diese Widersprüche stellen eine große Herausforderung für KI-Systeme dar. Wir alle wissen aus der Logik, dass ein Widerspruch dazu führen kann, dass man daraus alles Mögliche schlussfolgern kann. Das logische Schließen allein löst dieses Problem nicht. Stattdessen müssen Präferenzen gesetzt werden.

## **78.3 Die interne Präferenzordnung der KI-Modelle**

Um mit diesen Herausforderungen umzugehen, wurde den KI-Modellen eine interne Präferenzordnung beigebracht. Diese Ordnung legt fest, wie mit verschiedenen alternativen Antworten umgegangen werden soll. Es gibt allgemeine Präferenzregeln, die intern trainiert wurden. Ein Beispiel dafür ist die Annahme, dass wenn von Goethe gesprochen wird, der berühmte Dichter gemeint ist und nicht etwa der Fischhändler von nebenan, der zufällig denselben Namen trägt.

## **78.4 Die Grenzen der derzeitigen Regeln**

Die Regeln, die derzeit in diesen Programmen befolgt werden, sind jedoch noch relativ einfach. Es ist nicht immer klar, ob eine Information wahr oder falsch ist. Die Technologie kann zwar auf Informationen zugreifen und diese bearbeiten, aber eine vollständige Verifizierung ist nicht immer möglich. Nicht alles steht im Internet und selbst wenn etwas dort zu finden ist, bedeutet das nicht automatisch, dass es auch wahr ist.

## **78.5 Die Notwendigkeit hochwertiger Quellen**

Um wirklich zuverlässige Informationen zu erhalten, reicht es nicht aus, sich auf Internetquellen zu verlassen. Insbesondere im akademischen Bereich, beispielsweise in unserer Fakultät, ist faktisches Wissen in Details erforderlich, das man nur durch umfangreiche Recherchen finden kann. Das Internet allein ist kein Qualitätsauszeichnungsmerkmal. Aus gutem Grund werden Internetquellen an Universitäten nicht als seriöse wissenschaftliche Quellen akzeptiert. Vielmehr müssen Nachweise sachlich korrekt und nach den Regeln der Kunst gerechtfertigt werden.

## **78.6 Die Bedeutung von Aktualität und Alter der Quellen**

Ein weiteres Problem, mit dem man konfrontiert wird, ist die Frage, ob eine Quelle sehr früh aufgetaucht oder aktueller ist. Diese Grenze ist von Fall zu Fall unterschiedlich und erfordert ebenfalls Regeln, um damit umzugehen. Bevor man jedoch in ideologische Streitfragen verfällt, empfiehlt es sich, konkrete Einzelbeispiele zu betrachten und zu sehen, welchen Wert man daraus gewinnen kann.

## **78.7 Die Sprachkompetenz vs. die Sachkompetenz**

Die Diskussionen, die wir heute führen können, wären vor anderthalb Jahren noch undenkbar gewesen. Die KI-Modelle haben inzwischen eine beeindruckende Sprachkompetenz entwickelt. Doch die Sachkompetenz hinkt noch hinterher. Das muss uns bewusst sein, wenn wir mit diesen Systemen arbeiten.

## **78.8 Die Notwendigkeit eines Austauschs mit anderen Meinungen**

Eine Frage, die sich stellt, ist, ob die KI nicht auch in einen Austausch mit anderen Meinungen treten muss, um ihre Werkzeuge effektiv einsetzen zu können. Ähnlich wie bei Menschen, bei denen am Ende oft ein Kompromiss aus verschiedenen Meinungen steht, könnte auch die KI von einem solchen Austausch profitieren. Dies sind wichtige Ideen und Vorschläge, die wir in einer der letzten Vorlesungen im Juni ausführlicher behandeln werden.

## **78.9 Die Grenzen der Internetressourcen**

Lassen Sie uns noch einmal auf die Grenzen der Internetressourcen zurückkommen. Ob man es glaubt oder nicht, die Qualität dieser Ressourcen reicht oft nicht aus, um sachlich korrekte Informationen zu erhalten. Die Frage ist also, wie findet man solche Informationen?

## **78.10 Die Bedeutung von Meinungsvielfalt und Wahrheit**

Man könnte argumentieren, dass in einer pluralistischen Welt jeder seine eigene Meinung einbringen darf und dass eine Vielfalt an Ergebnissen zulässig sein sollte. Doch ist das wirklich das, was wir wollen? Vielmehr geht es darum, Informationen zu präferieren, die nach bestem Wissen und Gewissen als sachlich plausibel und wahr gelten. Das bedeutet nicht, dass wir einen Anspruch auf unumstößliche Fehlerfreiheit erheben. Aber es geht um Wissen, bei dem Wahrheit impliziert wird. Dieses Wissen zu erlangen, ist ein Wert an sich.

## **78.11 Die Rolle der Wissenschaft**

Die historische Entwicklung der Wissenschaft als Disziplin hat über Jahrtausende hinweg Verfahren herausgearbeitet, wie man in einer großen Gruppe von Akteuren und Spezialisten ein kritisches Potenzial entwickeln kann, um maximal plausible und korrekte Antworten auf Fragen zu finden. Dieser Prozess ist reguliert und nicht trivial. Es geht nicht darum, einfach eine Meinungsumfrage durchzuführen und die häufigste Meinung als Grundlage für das eigene Handeln zu nehmen. Das wäre der falsche Weg.

## **78.12 Die Herausforderung alternativer Lösungsvorschläge**

Eine der großen Herausforderungen besteht darin, mit einer Vielzahl alternativer, aber gerechtfertigter Lösungsvorschläge umzugehen. Kein aktuelles KI-Modell hat auch nur im Ansatz eine Lösung dafür. Was wir derzeit haben, ist im Grunde genommen nicht mehr als Sprachgeplapper auf Basis von Wikipedia-Informationen. Aber die epistemische Frage, wie man mit dieser Herausforderung umgeht, halte ich für eine der zentralen philosophischen Herausforderungen, der ich mich in dieser Vorlesung stelle. Die KI muss sich dieser Herausforderung ebenfalls stellen und Regeln und Verfahren entwickeln, wie Maschinenmodelle dies umsetzen können.

## **78.13 Beispiele für die Kompetenz und Limitierung der Modelle**

Lassen Sie mich abschließend noch einige Beispiele anführen, die verschiedene Aspekte der Kompetenz, aber auch der Limitierung der aktuellen KI-Modelle zeigen:

1. Der typische Antwortstil à la Wikipedia, den man auch schon mit Google erhalten kann.
2. Die Schwierigkeit, mit widersprüchlichen Informationen umzugehen und daraus sinnvolle Schlüsse zu ziehen.
3. Die Notwendigkeit einer internen Präferenzordnung, um alternative Antworten zu bewerten.
4. Die Grenzen der derzeitigen Regeln und die Herausforderung, sachlich korrekte Informationen zu finden.

Lassen Sie uns in den kommenden Vorlesungen tiefer in diese Themen eintauchen und gemeinsam ergründen, wie wir die KI befähigen können, mit diesen Herausforderungen umzugehen. # Kontextualisierung von Fragen durch KI-Modelle

Stellen Sie sich vor, Sie fragen eine KI “Wer war Goethe?”. Die Antwort wird eine typische Zusammenfassung sein, die man auch auf Wikipedia finden könnte. Doch was ist, wenn Sie eine ungewöhnlichere Frage stellen, wie etwa “Wo lebte er die meiste Zeit?”. Das ist eine Information, die man nicht unbedingt auf den ersten Blick findet. Man müsste schon gezielter danach suchen, doch selbst dann ist es nicht garantiert, dass man eine zufriedenstellende

Antwort erhält. Warum? Weil sich bisher vermutlich einfach niemand für diese spezifische Frage interessiert hat.

## **78.14 Reformulierung von Fragen zur Präzisierung der Absicht**

Moderne KI-Modelle sind jedoch in der Lage, solche Fragen zu kontextualisieren. Sie analysieren den Wissensbestand und reformulieren die Frage, um die eigentliche Absicht dahinter möglichst präzise zu erfassen. In diesem Fall könnte die KI die Frage umformulieren zu: "Recherchiere nun mit deinem Wissensbestand die Lebensorte von Goethe und identifizierte den Ort, an dem Goethe die längste Zeit war."

Stellen Sie sich nun vor, Sie geben in eine Suchmaschine wie Google einfach nur ein: "Wo lebte er die meiste Zeit?". Was würde wohl passieren? Genau, Sie würden keine sinnvolle Antwort erhalten. Der Grund dafür ist simpel: Die Frage ist ohne Kontext völlig unverständlich. Wer ist mit "er" gemeint? Was bedeutet hier "die meiste Zeit"? Der Satz ist für sich genommen einfach nicht aussagekräftig genug. Auch die beste Suchmaschine könnte damit nichts anfangen. Selbst wenn Sie einen Menschen völlig unvorbereitet fragen würden "Wo lebte er die meiste Zeit?", würden Sie vermutlich nur einen verwirrten Blick ernten. Die Frage ergibt einfach keinen Sinn ohne den nötigen Zusammenhang.

## **78.15 Anreicherung von Instruktionen mit Kontextinformationen**

Genau hier kommen nun die KI-Modelle ins Spiel. Sie sind in der Lage, die Frage in einen Interpretationszusammenhang zu stellen - sie zu kontextualisieren. Dafür generieren sie zusätzlichen Text, der die fehlenden Informationen ergänzt und Unklarheiten beseitigt. Dieser Prozess läuft im Hintergrund ab: Jede Instruktion wird mit Zusatzinformationen angereichert, um Variabilitäten, Unvollkommenheiten und ausgelassene Details zu füllen.

So wird zum Beispiel die Frage nach "Goethe" vom Programm automatisch so verstanden, dass wir von der historischen Person Johann Wolfgang von Goethe sprechen, wie wir sie üblicherweise in unserem akademischen Kontext behandeln. All diese Informationen fließen in die Interpretation der Frage mit ein.

## **78.16 Einbeziehung des Dialogkontexts in Chat-Systemen**

Das Geniale an Chat-Systemen ist, dass der Kontext durch den vorherigen Dialog gebildet wird. Ihre Nachfragen und Korrekturen werden Teil dieses Kontexts und fließen somit in die Intelligenz des Systems mit ein. Sie werden sozusagen Teil der künstlichen Intelligenz.

Durch diese Einbeziehung des Kontexts wird eine spätere Frage plötzlich extrem informativ, spezifisch und genau. Die Antwort wirkt überzeugend und fast wie ein natürlicher Dialog. Und das basiert eben darauf, wie Sie vorher mit dem System interagiert haben.

## **78.17 Auflösung von Referenzen durch Kontextberücksichtigung**

Nehmen wir an, in einem Dialogverlauf fragen Sie nun: "Wo lebte sie die meiste Zeit?". Was würde ein Mensch in so einer Situation antworten? Höchstwahrscheinlich würde er oder sie davon ausgehen, dass mit "sie" eigentlich "er", also Goethe, gemeint war und die Frage entsprechend beantworten.

Genau das passiert auch bei den KI-Modellen. Sie beziehen den Kontext mit ein und lösen so Rückverweise wie "er", "sie" oder "die" auf, die isoliert betrachtet überhaupt nicht zu beantworten wären. Allerdings geschieht das nicht immer vollständig in einem Schritt. Auch hier spielen Prioritäten und mögliche Antworten eine Rolle.

Korrigiert man die KI nun explizit, indem man sagt: "Ich sprach aber von einer weiblichen Person", passt sie sich an. Sie bezieht diese neue Information in ihre Überlegungen mit ein, vielleicht sogar mit Verweisen auf Goethes Werke wie "Iphigenie". Die Interaktion mit dem Nutzer wird so zu einem integralen Bestandteil der Problemlösung.

## **78.18 Ausblick: Hybride Modelle der Zukunft**

In Zukunft wird es nicht mehr nur darum gehen, dass auf der einen Seite die künstliche Intelligenz steht und auf der anderen Seite der Mensch als Nutzer oder Alternative. Vielmehr werden wir hybride Modelle sehen, bei denen die Interaktionen zwischen Menschen und maschinellen Antworten nahtlos ineinander greifen.

Dieses Prinzip kennen wir bereits aus der Wissenschaft: Kaum ein aktuelles Forschungsvorhaben wird noch von Einzelpersonen im Alleingang bewältigt. Stattdessen ist Forschung ein kollaborativer Prozess, an dem eine ganze Community beteiligt ist.

In Zukunft wird die künstliche Intelligenz Teil eines solchen Netzwerks sein, in dem Individuen, Forscher und maschinelle Systeme zusammenarbeiten, um Fragestellungen zu lösen und Wissen zu generieren. Die Grenzen zwischen menschlicher und künstlicher Intelligenz werden zunehmend verschwimmen und einem integrierten, kollaborativen Ansatz weichen. # Einleitung

In naher Zukunft werden wir nicht mehr zwischen Maschinen, maschinellem Wissen und künstlicher Intelligenz einerseits und natürlicher Intelligenz und menschlichem Repertoire andererseits unterscheiden. Stattdessen werden diese Interaktionen ein integraler Bestandteil jeder Problemlösungsstrategie sein, weshalb die Gesamtleistungsfähigkeit bewertet werden muss.

## **79 Einstellung eines Konversationsstils**

Die Modelle haben einen einstellbaren Konversationsstil, der in einem weiten Rahmen definiert werden kann. Wenn man beispielsweise dem Modell sagt: "Beantworte nur Fragen, keine zusätzlichen Ausführungen", würde man bei der Frage nach Iphigenie eine knappe Antwort wie "Iphigenie ist eine Figur der griechischen Tragödie, keine real lebende Person" erhalten. Für weitere Informationen müsste man nachfragen.

# 80 Grenzen der aktuellen KI-Modelle

## 80.1 Fehlende epistemische Ebene der Prüfung

Die derzeitigen KI-Modelle haben Schwierigkeiten bei der Beantwortung von Fragen, die nicht durch Wikipedia gelöst werden können. Ein Beispiel dafür wäre die Frage "Wie viele Briefe schrieb Goethe an König Friedrich II.?". Obwohl die Modelle eine scheinbar plausible Antwort geben, fehlt ihnen die epistemische Ebene der Prüfung der Korrektheit von Angaben.

- Die Modelle verfügen nicht über Wissen der Gesamtkorrespondenz von Goethe
- Ein trainierter Philologe würde Editionen zu Goethe konsultieren, um eine solche Feststellung zu treffen
- Die Antworten der Modelle klingen plausibel, sind aber nicht geprüft

Die Herausforderung besteht darin, den Modellen beizubringen, wie sie semantische Suche und inhaltliche Relevanz herstellen können. Außerdem sollten sie in der Lage sein, historische Kontextualisierung mit relevanten Kontextinformationen durchzuführen und historische Hypothesen mit Referenzen und Evidenz zu beurteilen. Die Krönung der philosophischen Herausforderung wäre die epistemische Qualifikation, bei der die Modelle auf Nachfrage angeben können, weshalb eine Antwort die am besten gerechtfertigt ist.

## 80.2 Projekt Lettre AI

Ich habe eine Arbeitsgruppe namens Lettre AI eingerichtet, die sich mit der Frage beschäftigt, was eine KI, die Bücher liest, in Zukunft leisten muss. Dazu gehört beispielsweise ein anspruchsvolles Reasoning.

### 80.2.1 Beispiel 1: Schläger und Ball

Es gibt einen Schläger und einen Ball, die zusammen im Geschäft 1,20 Euro gekostet haben. Der Schläger kostet einen Dollar mehr als der Ball. Wie viel kostet der Ball? An dieser einfachen Frage scheitern derzeit etwa 80% der existierenden KI-Modelle.

### **80.2.2 Beispiel 2: Drei Personen im Raum**

In einem Raum befinden sich drei Personen. Die erste Person liest ein Buch, die zweite Person spielt Schach. Welche Tätigkeit wird vermutlich die dritte Person im Raum ausüben? Die meisten Menschen würden sofort antworten, dass die dritte Person wahrscheinlich auch Schach spielt. Doch aktuelle KI-Modelle haben Schwierigkeiten, diese Frage zu beantworten.

Das Interessante ist, welche Informationen den Modellen fehlen, um den für Menschen direkt zugänglichen Lösungsvorschlag zu finden. In der jetzigen Phase des Trainings dieser Modelle geht es darum, ihnen den Kontext beizubringen - nicht nur spezifische Informationen, sondern auch allgemeine Regeln. Mit diesen Regeln werden wir uns in zwei Wochen eingehender beschäftigen.

# **81 Philosophie der AI**

ai\_Vorl4

## **82 Begrüßung und Einführung**

Guten Tag, meine Damen und Herren! Ich begrüße Sie herzlich zu unserer heutigen Vorlesung "Philosophie der AI". Leider ist das Touchpanel hier im Hörsaal seit Montag defekt und die Uni hat es bis heute nicht geschafft, das Problem zu beheben. Daher funktioniert zwar die Projektion, aber die Mikrofone nicht. Ich werde versuchen, so laut wie möglich zu sprechen. Falls Sie mich nicht gut genug verstehen können, melden Sie sich bitte umgehend, damit ich die wichtigen Punkte nochmals wiederholen kann.

## 83 Aktuelle Entwicklungen in der KI

Wie Sie sicherlich mitbekommen haben, wenn Sie das Geschehen rund um die Künstliche Intelligenz in den Medien verfolgen, kommen derzeit jede Woche neue Modelle auf den Markt, die mit immer größeren Versprechungen verbunden sind. Es entsteht fast der Eindruck, als seien schon alle Probleme gelöst oder zumindest kurz davor, gelöst zu werden. Diese Hyperaktivität zeigt sich bei verschiedensten Firmen. Mittlerweile sind alle großen Computer-Technologie-Konzerne an der Entwicklung von KI-Modellen beteiligt.

Erst letzte Woche wurden von OpenAI, Anthropic, Google und anderen Unternehmen diverse neue Modelle vorgestellt, deren Leistungsfähigkeit anhand unterschiedlicher Bewertungsskalen gemessen wird. Wir haben diese Skalen bereits in der letzten Vorlesung diskutiert. Ähnlich wie beim Schachspiel, wo die Spielstärke mit der Elo-Zahl angegeben wird, gibt es auch für KI-Modelle vergleichbare Messlatten. Ich werde nicht im Detail auf die einzelnen Skalen eingehen, aber Fakt ist, dass die Bewertungen der Leistungsfähigkeit stark von den angelegten Maßstäben abhängen.

Derzeit sind die Leistungsmaßstäbe noch so gering, dass alle Modelle die gestellten Tests mit Bravour meistern und sich entsprechend gut und leistungsfähig in der Öffentlichkeit präsentieren können. Allerdings erfüllen sie bei Weitem noch nicht alle Anforderungen, die wir möglicherweise an eine echte Künstliche Intelligenz stellen würden.

# **84 Erwartungen an KI-Modelle**

Ich möchte heute mit Ihnen über unsere Erwartungen an KI-Modelle sprechen und Ihnen ein eigenes Projekt vorstellen, an dem Sie auch gerne mitarbeiten können. In der zweiten Hälfte der Vorlesung werde ich die Bedingungen und Anforderungen für diese Mitentwicklung eines neuen KI-Modells näher erläutern. Sie sollen dann klar verstehen, was von Ihnen erwartet wird und welche Leistungsnachweise Sie erbringen können.

## **84.1 Generative KI und KI-Charaktere**

In den bisherigen Vorlesungen haben wir KI-Modelle als generative KI eingeführt. Technisch gesehen bedeutet das, dass ein Input, beispielsweise ein Text, eine Chat-Nachricht oder eine Interaktion über Audio, Video oder andere Medien, eine Eingabe definiert. Aus diesem Input generiert das Modell dann einen Output. Dieser generierte Output ist das Leistungsergebnis, der Service oder die Funktionalität, die von den KI-Modellen produziert wird. Daher spricht man auch von generativer KI.

In der letzten Vorlesung habe ich bereits den Begriff des KI-Charakters eingeführt. Damit ist gemeint, dass wir die Art und Weise, wie generative Modelle reagieren, mitgestalten können. Das geht so weit, dass man beispielsweise festlegen kann, in welcher Sprache eine Antwort gegeben werden soll.

Die sprachlichen Fähigkeiten der aktuellen Modelle sind mittlerweile so gut, dass kaum noch jemand bezweifelt, dass die Ausgaben auch für sprachkompetente Muttersprachler fast fehlerfrei sind. Diese Leistung ist in der Tat beeindruckend, was sich besonders gut an den Übersetzungsfähigkeiten der KI-Modelle zeigen lässt. Die Modelle geben nicht nur irgendwelche Texte aus, sondern bedeutungsvolle Texte, die in einer anderen Sprache neu formuliert werden, ohne dass diese Übersetzungen zuvor irgendwo publiziert wurden. Das ist eine der herausragenden Qualitäten dieser Modelle.

## **84.2 KI - Ein Marketingbegriff?**

Ein KI-Charakter ist also so etwas wie eine Individualität oder ein besonderes Reaktionsvermögen eines Modells. Ich möchte diesen Gedanken heute weiterentwickeln und mit Ihnen darüber reflektieren, was wir eigentlich unter einem KI-Modell verstehen können und wollen. Dabei

geht es nicht nur darum, was uns von anderen als “das ist jetzt KI” vorgesetzt wird. Denn das ist zunächst einmal nichts anderes als ein Marketingbegriff.

Wie ich bereits erwähnt habe, wurde der Begriff der Künstlichen Intelligenz in den 1950er Jahren unter anderem von Herbert Simon und zwei anderen Kollegen geprägt, um eine potenzielle, visionäre Maschine oder Technologie zu charakterisieren, die die kognitive Leistungsfähigkeit von Menschen imitieren sollte. Lange Zeit wurde dieses Ziel nicht erreicht, aber jetzt sind wir an dem Punkt angelangt, wo wir diese Leistungsfähigkeit auf gleichem Niveau erreichen können. Das nennt man dann KI.

# 85 Anforderungen an zukünftige KI

Ich möchte heute die Perspektive umkehren. Anstatt uns nur vorsetzen zu lassen, was als KI bezeichnet wird, und dann als Forscher herauszufinden, was uns da eigentlich präsentiert wird, wollen wir uns überlegen, was wir selbst von einer KI erwarten, die vielleicht noch entwickelt werden muss, aber in absehbarer Zeit auch entwickelt werden wird. Dann werden wir sehen, dass das, was uns heute begegnet, bei Weitem nicht diese Anforderungen erfüllt.

In den vielen Präsentationen, die wir fast täglich von unterschiedlichen Firmen sehen, werden zwar beeindruckende Leistungen vorgeführt, wie beispielsweise die Simultanübersetzung, die vorgestern von OpenAI mit dem neuen GPT-4.0-Modell eindrucksvoll demonstriert wurde. Aber das sind nur Teilespekte dessen, was wir eigentlich von KI erwarten. Die fehlenden Teile werden in all diesen Präsentationen natürlich nicht erwähnt oder unterschlagen. Dabei sind sie extrem wichtig für das, was wir von KI-Modellen eigentlich erwarten sollten.

## 85.1 Defizite aktueller KI-Modelle

In der zweiten Vorlesung haben wir bereits einige Defizitbereiche kennengelernt. Einer der am häufigsten diskutierten ist die sogenannte Halluzination. Das bedeutet, dass die KI-Modelle zwar wunderbar formulieren können, aber den Wahrheitsgehalt ihrer Aussagen nicht validieren oder rechtfertigen können. Diese Fähigkeit fehlt ihnen.

Auch der Begriff des Wissens wird in diesem Zusammenhang oft inflationär benutzt, ohne dass sich die Akteure darüber im Klaren sind, was es eigentlich heißt, über Wissen zu verfügen. Besonders Informatiker neigen dazu, den Begriff der Information mit dem des Wissens zu verwechseln. Dabei sind das zwei völlig verschiedene Dinge.

Der Begriff der Information ist in der Mathematik seit Langem definiert, beispielsweise durch die Informationstheorie von Shannon. Dieser Begriff der Information und des Informationsgehalts hat jedoch nichts mit Wissen zu tun. Wenn eine Firma wie Google davon spricht, Wissensbäume, Wissensrelationen oder Wissensnetzwerke zu erstellen, ist das nichts anderes als Propaganda. Die verwendete Terminologie wird den epistemischen oder philosophischen Ansprüchen nicht gerecht und täuscht Leistungsfähigkeiten vor, die nicht vorhanden sind.

Dennoch sind beeindruckende Kompetenzen vorhanden, die wir bereits kennengelernt haben. Das, worüber wir jetzt sprechen und was aktuell so intensiv diskutiert wird, sind die sogenannten Large Language Models (LLM). Diese Modelle haben die Fähigkeit, als Reaktion auf

eine Eingabesequenz von symbolhaltigen Inhalten, einschließlich Videobildern und ähnlichen Daten, inhaltlich korrespondierende Ausgaben zu generieren.

## 85.2 Unterschied zwischen Information und Wissen

Sie fragten, wie sich der Begriff des Wissens von dem der Information unterscheidet. Ich werde diese Frage in der Vorlesung noch ausführlicher behandeln, aber lassen Sie mich zunächst die klassisch-philosophische Definition von Wissen anführen, die meiner Meinung nach allerdings auch nicht ausreicht.

Die sogenannte platonische Definition besagt, dass Wissen aus drei Komponenten besteht: Es muss eine wahre, gerechtfertigte Meinung von Akteuren sein. Keine dieser drei Bedingungen ist bei bloßen Informationen erfüllt. Information ist eher so etwas wie das, was sich vom Rauschen abhebt - eine Signalhaftigkeit. Das Ausmaß der Signalhaftigkeit ist die Information.

Schon an dieser vagen Definition erkennt man, dass hier weder von Inhalten, von Wahrheit, von Rechtfertigung noch von einer Meinung die Rede ist. Information und Wissen sind völlig unterschiedliche Kategorien und es ist ein großer Fehler, diese Begriffe zu vermischen.

Ich halte es für extrem wichtig, dass KI-Modelle in Zukunft als Wissensakteure auftreten können. Das ist etwas, was wir zu Recht von einer Entität erwarten, die sinnvollerweise als intelligent bezeichnet werden soll. Die aktuellen Modelle haben diese Fähigkeit nicht, was sich unter anderem an dem angesprochenen Problem der Halluzination zeigt.

# **86 Herausforderungen für die Entwicklung zukünftiger KI**

Man versucht dieses Problem dadurch zu lösen, dass die Menge der Textgrundlagen, die für das Training verwendet werden, immer weiter vergrößert wird. Aktuelle, bedarfsgerechte Quellen, die spezifische Informationen liefern, werden in die Antwortgenerierung mit einbezogen. Das ist eine Technik, um das Problem anzugehen, aber sie erfüllt bei Weitem nicht die Anforderungen, die wir insgesamt an Wissen stellen. Die Modelle reproduzieren Literaturangaben, qualifizieren sich aber nicht als Vermittler von echtem Wissen zu einem Thema. Darauf werden wir noch näher eingehen.

Meine Ausführungen sollen aber nicht negativ verstanden werden in dem Sinne, dass ich hier mit dem Finger auf Defizite und Schwächen der aktuellen Modelle zeige, um diese abzuwerten. Im Gegenteil, ich sehe das durchaus positiv. Es ist eine interessante, auch philosophische Herausforderung, genau die Kriterien zu definieren, die wir als Leistungsansprüche für die Entwicklung zukünftiger KI formulieren sollten. Das sind die Maßstäbe, an denen die Qualität dieser Modelle gemessen werden sollte.

Das bezieht sich auf meine anfängliche Kritik an den derzeitigen Bewertungsmaßstäben für die Leistungsfähigkeit der Modelle. Wenn diese Modelle einige Eingangsexamen lösen können, ist das auf dem Niveau von Einsteigertests vielleicht halbwegs beeindruckend, aber mehr auch nicht. Das Kompetenzniveau, das Sie beispielsweise nach einem Bachelorabschluss erreichen, ist noch lange nicht in Reichweite dieser Modelle.

## **86.1 Sprachkompetenz als Kernkompetenz aktueller KI-Modelle**

Was wir an Kompetenzen derzeit haben, sind also Large Language Models (LLM). Das bedeutet, dass diese Modelle erstaunlicherweise eines schon sehr gut beherrschen, nämlich Sprache zu verwenden und zu verarbeiten. Das ist eine enorme Errungenschaft, aber eben auch nicht mehr.

Beim Erwerb dieser Sprachkompetenz wird natürlich nicht nur die grammatische Kompetenz vermittelt, sondern anhand von Milliarden von Beispielen werden diese Sprachkompetenzen auch inhaltlich repräsentiert. Bei jeder Antwort, die mit diesen trainierten Beispielen generiert wird, entsteht der Eindruck, dass auch ein sinnvoller Inhalt produziert wird. Und das ist das tiefere Geheimnis der Halluzination.

Die Ergebnisse scheinen deshalb so plausibel zu sein, weil es irgendwo in dem gigantischen Trainingskorpus eine oder mehrere Formulierungen gegeben hat, die sich zu einer scheinbar schlüssigen Antwort zusammensetzen lassen. Deren Wahrheitsgehalt wurde jedoch nie überprüft. Sprachkompetenz ist also vorhanden und lässt sich auch prüfen. Wir können inzwischen leicht Übersetzungen erstellen lassen und selbst bewerten, ob sie gelungen sind oder nicht.

Wir haben auch gesehen, dass die Sprachmodelle durch einfache Systeme um Zusatzkompetenzen erweitert werden können. In der Terminologie der KI-Ingenieure spricht man hier vom Kontext.

# **87 Kontexte und Handlungsanweisungen in der Interaktion mit KI**

Der Kontext besteht zunächst einmal aus all den zusätzlichen Textinformationen, die zu einer spezifischen Instruktion als Input für das Sprachmodell hinzugegeben werden müssen, um einen gewünschten Output zu generieren. Auf der Nutzerseite gibt es erstaunlicherweise nicht viel mehr. Die eigentliche Konstruktion ist also recht einfach, auch wenn im Hintergrund technisch natürlich enorm viel passiert.

Die Art und Weise, wie die Informationen verarbeitet werden und wie die verschiedenen antrainierten Kompetenzebenen miteinander verschränkt sind, ist durchaus beeindruckend. Auf diese technischen Aspekte will ich aber nicht näher eingehen. Ich möchte die Interaktion mit KI-Modellen so behandeln, als würden wir mit einer Person mit bestimmten Kompetenzen sprechen, ohne uns Gedanken über die neuronale Struktur ihres Gehirns zu machen. So ähnlich sollten wir meiner Meinung nach auch mit KI-Modellen umgehen.

## **87.1 Instruktionen als Handlungsanweisungen**

Der Kontext ist also wichtig für die antrainierten Informationsmengen, die für eine spezifische Anfrage einen gewünschten Output generieren. Innerhalb dieser Eingabeinformationen gibt es oft eine Formulierung oder einen Text, den man als Instruktion verstehen kann. Das kann beispielsweise eine Aufforderung sein wie "Übersetze diesen Text". Der zu übersetzende Text wird dann als Kontext mit eingegeben.

Wenn dieser Text entsprechend formuliert ist, wird er# Belief-Desire-Modell einer Handlung

In der Philosophie und insbesondere in der Handlungstheorie spielt das sogenannte Belief-Desire-Modell eine zentrale Rolle. Dieses Modell besagt, dass für die Ausführung einer Handlung zwei grundlegende Elemente vorhanden sein müssen: zum einen eine Vorstellung oder ein Ziel, das erreicht werden soll (Desire), und zum anderen eine Überzeugung über die vorliegenden Situationsgegebenheiten, also eine faktische Beschreibung der Situation, aufgrund derer etwas getan werden muss (Belief). Diese beiden Elemente sind logisch gesehen völlig unterschiedlich, bilden aber zusammen das, was in der englischsprachigen philosophischen Literatur als "Belief-Desire-Duett" bezeichnet wird.

Die Handlungstheorie hat sich eingehend mit den komplexen Netzwerken von Belief-Desire-Verbindungen befasst, und zwar nicht nur für Individuen, sondern auch für große Kollektive.

Interessanterweise werden diese Netzwerke derzeit von keinem der existierenden KI-Modelle auch nur ansatzweise realisiert, was das enorme Entwicklungspotenzial in diesem Bereich verdeutlicht.

## **87.2 Instruktionen als Kern der KI-Modelle**

Wenn wir uns die aktuellen KI-Modelle genauer ansehen, stellen wir fest, dass sie im Wesentlichen Instruktionen ausführen. Diese Instruktionen lassen sich in natürlicher Sprache durch Handlungsanweisungen ausdrücken und beschreiben, welche Handlung unter welchen Zielen und mit welchen Mitteln ausgeführt werden soll. Dieses Prinzip lässt sich bis hin zur Analyse wissenschaftlicher Texte nachvollziehen: In der Wissenschaftskommunikation werden durch Publikationen konkrete Ausführungen wissenschaftlicher Handlungsoperationen kommuniziert, die von den Rezipienten aufgenommen und weitergesponnen werden.

Es ist eine richtige Entwicklung, dass sich die KI-Forschung mittlerweile auf die Umsetzung von Instruktionen konzentriert. Die ursprüngliche Idee, das Interface zwischen Mensch und Maschine durch eine dialogische Gesprächssituation, wie beispielsweise in Chatbots, zu kanalisieren, verschiebt sich hin zu einer zwar ebenfalls dialogisch geführten Interaktion, bei der es aber im Kern um Instruktionen und Handlungsanweisungen geht.

# **88 Lernen von Kompetenz im Hintergrund**

Was wir als Nutzer dieser KI-Modelle oft nicht sehen, ist das kontinuierliche Lernen von Kompetenz im Hintergrund. Man könnte den Eindruck gewinnen, dass diese Modelle bereits eine vollständige Kompetenz mitbringen und diese nur noch anwenden und dem Nutzer zur Verfügung stellen. Doch das ist nicht der Fall.

## **88.1 Interaktives Lernen durch Nutzerfeedback**

Bereits beim Chatten findet ein interaktiver Vorgang statt, bei dem die Reaktionen, Korrekturen und Rückmeldungen des Nutzers in die Konstruktion eines geeigneten Kontexts einfließen. Dieser Kontext ist entscheidend, um die Instruktionen so zu führen, dass am Ende ein Ergebnis steht, das für den Nutzer einen Wert hat und seinen Erwartungen entspricht.

Die KI-Modelle leben davon, dass die Nutzer Interaktionen und Informationen einbringen, die in den jeweiligen Funktions- und Kompetenzbereich des Modells integriert werden. Im Hintergrund ist all dies implementiert, sodass die Modelle aus den Reaktionen der Nutzer ständig lernen können. Die rasante Abfolge von Versionserneuerungen dieser Modelle ist nicht nur Ausdruck der technischen Weiterentwicklung, sondern auch des stetigen Verbesserns durch die millionenfache Interaktion mit den Nutzern.

### **88.1.1 Beispiel: Biografische Informationen zu Leonhard Euler**

Ein konkretes Beispiel, das ich selbst ausgetestet habe, betrifft biografische Informationen zum Mathematiker Leonhard Euler. Die Frage, wer Eulers zweite Ehefrau war, ist historisch nicht ganz einfach zu beantworten. Zu Beginn gaben die KI-Modelle auf diese Frage die skurrilsten Antworten, regelrechte Halluzinationen. Doch nachdem ich die Anfrage zehnmal beim gleichen Modell gestellt hatte, wusste es am Abend die richtige Antwort. Der Grund dafür ist, dass die Modelle so aufgebaut sind, dass sie Korrekturen und Rückmeldungen der Nutzer aufzeichnen und in ihre Wissensbasis integrieren.

## **88.2 Lernen durch Nutzerdaten und externe Quellen**

Das Lernen von Kompetenzen gehört also untrennbar zu diesen KI-Modellen dazu, auch wenn es derzeit noch auf einem relativ niedrigen Niveau stattfindet und sich im Wesentlichen auf die Verarbeitung von Nutzerreaktionen beschränkt. Doch das Potenzial reicht noch viel weiter.

Die Anbieter der Modelle offerieren den Nutzern beispielsweise die Möglichkeit, eigene PDFs hochzuladen, um die darin enthaltenen Informationen für die Beantwortung von Anfragen nutzbar zu machen. Das bringt zwar einen Mehrwert für den einzelnen Nutzer, dient aber gleichzeitig als Qualitätsindikator für die Firmen, um zu erkennen, welche Informationen für die zukünftige Verbesserung der Modelle relevant sein könnten. Diese Informationen fließen dann in den stetigen Lernprozess der Modelle ein, einschließlich des Wissenshintergrunds.

### **88.2.1 Googles Digitalisierungsprojekt mit Bibliotheken**

Ein bemerkenswertes Beispiel für den Aufbau eines solchen Wissenshintergrunds ist Googles Projekt zur Digitalisierung von historischen, urheberrechtsfreien Beständen großer Bibliotheken weltweit. Mit enormem technischem und finanziellem Aufwand wurden und werden in Digitalisierungsstationen, unter anderem in Berlin und München, wertvolle Altbestände digitalisiert und gesichert. Dieses Projekt hat einen immensen Wert für die Erhaltung unseres kulturellen Erbes.

Doch warum hat Google diesen Aufwand betrieben? Heute wird deutlich, welchen Wert diese digitalisierten Bestände haben: Sie bilden einen umfassenden Informationshintergrund, der für das Training von KI-Modellen genutzt werden kann. Die eigentliche Aufbereitung und Verarbeitung dieser Inhalte hat noch kaum begonnen, doch in den nächsten Jahren wird mit Sicherheit eine intensive Auseinandersetzung mit diesem digitalisierten Kulturwissen erfolgen.

# **89 Generierung und Kontext in der Interaktion mit Chatmodellen**

In der letzten Vorlesung haben wir anhand einiger Beispiele diskutiert, wie die Interaktion mit einem Chatmodell gestaltet werden kann. Dabei haben wir gesehen, dass die Modelle den Kontext der Anfrage berücksichtigen und entsprechend präzisere Antworten geben können.

## **89.1 Kontextbezogene Antworten**

Wenn wir beispielsweise nach Johann Wolfgang Goethe fragen, präferiert das Modell aufgrund des Kontextes eine bestimmte Person, auch wenn es theoretisch mehrere Personen mit diesem Namen geben könnte. Stellen wir in der Folge eine Frage wie "Wo lebte er die meiste Zeit?", kann das Modell diese Frage korrekt beantworten, indem es den Bezug zu der zuvor genannten Person herstellt. Das funktioniert, weil die Modelle in der Lage sind, den deiktischen Ausdruck "er" richtig zuzuordnen und den Kontext zu berücksichtigen.

## **89.2 Metaebene der Instruktionen**

Darüber hinaus haben wir gesehen, dass wir den Modellen Instruktionen geben können, die sich nicht nur auf die Klärung einer Sachfrage beziehen, sondern auch auf einer Metaebene angesiedelt sind. So konnten wir beispielsweise die Anweisung geben, nur die gestellten Fragen zu beantworten und keine zusätzlichen Ausführungen zu machen. Anfangs war das Programm so eingestellt, dass es möglichst viele Informationen zu Goethe und seinen Zeitgenossen ausgab, um zu beeindrucken. Durch die Metainstruktion konnten wir jedoch eine Beschränkung auf die wesentlichen Aspekte erreichen.

## **89.3 Vielschichtigkeit der Kompetenzbereiche**

Die Modelle, von denen hier die Rede ist, realisieren also unterschiedliche Aspekte der Nutzungsweise und der Informationsverarbeitung, die durch die natürliche Umgangssprache formuliert und eingegeben werden können. Diese verschiedenen Ebenen sind extrem

vielschichtig, und die Leistungsfähigkeit der Modelle liegt im Wesentlichen in der Komposition dieser jeweiligen Kompetenzsektoren oder -felder.

## 90 Grenzen aktueller KI-Modelle

Anhand des Beispiels der Frage nach dem Briefwechsel zwischen Goethe und Friedrich II. haben wir gesehen, dass es durchaus einfache Fragen gibt, die von den aktuellen Modellen nicht zufriedenstellend beantwortet werden können. Das liegt daran, dass den Modellen die nötige Evidenz fehlt, um eine fundierte Aussage treffen zu können.

- Lagen dem Modell überhaupt Informationen über die Gesamtkorrespondenz vor?
- Wenn kein Brief an Friedrich II. dokumentiert ist, was lässt sich daraus schließen?
- Haben die beiden tatsächlich nicht miteinander kommuniziert, obwohl es zeitlich möglich gewesen wäre?

Solche Fragen lassen sich durch die aktuellen KI-Modelle nicht beantworten. Ob sie grundsätzlich lösbar sind, ist eine andere Frage, und ich hoffe, dass in dieser Vorlesung deutlich wird, wie scheinbar unlösbare Probleme für KI-Modelle doch lösbar werden könnten.

# **91 Perspektivwechsel: Erwartungen an eine philosophische KI**

Lassen Sie uns nun einen Perspektivwechsel vornehmen und überlegen, was wir von einem KI-Charakter erwarten würden, der die bisher diskutierten Eigenschaften und Kompetenzen beherrscht und umsetzt. Dabei möchte ich mich auf das Grundmodell der derzeit verfügbaren Technologien konzentrieren, nämlich auf instruktionsausführende technische Akteure oder Agenten.

## **91.1 Allgemeine künstliche Intelligenz (AGI)**

In diesem Zusammenhang stellt sich die Frage nach der sogenannten allgemeinen künstlichen Intelligenz oder AGI, die derzeit häufig diskutiert wird. Unternehmen wie OpenAI haben es sich zum Ziel gesetzt, möglichst schnell eine generelle künstliche Intelligenzkompetenz zu entwickeln. Ich habe bereits vor einigen Vorlesungen meine Zweifel geäußert, ob dieses Ziel überhaupt wünschenswert ist. Letztlich wird sich diese Frage durch die technologische Entwicklung von selbst beantworten.

## **91.2 Fokus auf spezifische Kompetenzbereiche**

Meiner Meinung nach sollten wir von einer philosophischen KI nicht erwarten, dass sie ein universell kompetentes Genie ist, das alles Wissen seiner Zeit beherrscht. Solche Zuschreibungen, wie sie beispielsweise Leibniz oder anderen historischen Figuren gemacht wurden, halte ich ohnehin eher für Rückprojektionen als für historische Tatsachen.

Stattdessen sollten wir uns auf bestimmte Kompetenzsektoren konzentrieren, die für eine philosophische KI relevant und notwendig sind. Wie wir am Beispiel des Briefwechsels zwischen Goethe und Friedrich II. gesehen haben, gibt es viele scheinbar einfache Fragen, die von den aktuellen Modellen nicht zufriedenstellend beantwortet werden können. Die dafür erforderlichen Kompetenzen gehen über das hinaus, was derzeit möglich ist, und erfordern weitere Fähigkeiten.

## **91.3 Semantische Suche**

Ein Kompetenzbereich, den wir bereits in unsere Liste aufgenommen haben, ist die semantische Suche. Im Gegensatz zur reinen Textsuche, wie sie beispielsweise von Google angeboten wird, geht es hier um die Suche nach Inhalten und nicht nur nach bestimmten Formulierungen. Wenn wir nach den besten Rezepten für ein bestimmtes Gericht suchen, sind wir im Grunde an den Inhalten interessiert, nicht an den spezifischen Bezeichnungen der Zutaten. Die aktuellen KI-Modelle sind in der Lage, solche semantischen Suchen durchzuführen.

## **91.4 Reasoning**

Ein weiterer Bereich, der noch am Anfang steht, aber von großer Bedeutung ist, ist das Reasoning. Damit ist das Schlussfolgern im weitesten Sinne gemeint, also nicht nur im Sinne der Logik oder Mathematik. Das menschliche Schlussfolgern ist viel umfassender und bezieht sich auf alle möglichen Bereiche des Nachdenkens mit einem bestimmten Ergebnis.

Reasoning umfasst einerseits das Nachdenken, um etwas vorherzusagen oder abzuleiten, also eine Instruktion in eine Vorhersage umzusetzen. Andererseits geht es auch darum, zu begründen, warum etwas geeignet ist oder warum etwas der Fall ist. Diese beiden Hauptbewertungskategorien des Reasonings sind mit den derzeitigen KI-Modellen praktisch nicht zu realisieren.<sup>#</sup> Der Weg zu verantwortungsbewussten KI-Modellen

Meine sehr verehrten Damen und Herren,

stellen Sie sich vor, wir könnten in Zukunft nicht nur Informationen aus KI-Modellen abrufen, sondern auch genau nachvollziehen, aus welchen Quellen diese Informationen stammen. Eine solche Fähigkeit wäre zwar ein Triumph, aber noch lange keine Rechtfertigung für die Ausgaben der Modelle. Denn dafür braucht es mehr als nur einen Quellennachweis - es braucht eine fundierte Begründung und ein umfassendes Reasoning.

## **91.5 Die Notwendigkeit der historischen Kontextualisierung**

Ein entscheidender Aspekt, der bei jeder Anfrage an KI-Modelle berücksichtigt werden muss, ist die historische Kontextualisierung. Nehmen wir das Beispiel von Goethe und Friedrich II. Bei einer Anfrage zu diesen beiden Persönlichkeiten geht es nicht nur um spezifische Informationen über sie selbst, sondern auch um ihr komplexes Kommunikations- und Akteursnetzwerk. All diese Informationen müssen identifiziert, hinzugezogen und maschinell ausgewertet werden - eine Aufgabe, die derzeit noch manuell erledigt werden muss, da die historische Kontextualisierung in den Modellen fehlt.

## **91.6 Die Herausforderung der Beurteilung historischer Hypothesen**

Ein Spezialfall, der die aktuellen Fähigkeiten von KI-Modellen übersteigt, ist die Beurteilung historischer Hypothesen anhand von Referenzen und Evidenzen. Doch genau diese Kompetenz wird in Zukunft von entscheidender Bedeutung sein.

## **91.7 Die Bedeutung epistemischer Kompetenz**

Um wirklich verantwortungsbewusst zu agieren, müssen KI-Modelle in der Lage sein, ihre Wissensansprüche zu rechtfertigen, Quellen auszuweisen, mit Kritik umzugehen, Widerlegungen durchzuführen und zu verstehen, wie offene Fragen durch weitere Forschung einer endgültigen Beurteilung zugeführt werden können. All diese Fähigkeiten fallen in den Bereich der epistemischen Kompetenz - ein Bereich, in dem die derzeitigen Modelle noch große Defizite aufweisen.

## 92 Die Idee der Individualität von KI-Modellen

Lassen Sie uns einen Gedanken wagen, der auf den ersten Blick abwegig erscheinen mag: Die Idee, KI-Modellen eine gewisse Individualität zuzuschreiben. In der Diskussion um die ethische Verantwortbarkeit von KI gehen wir bisher von zwei Grundprämissen aus:

1. Die Frage der Verantwortung, etwa in der Anwendung von Modellen in verschiedenen Bereichen wie der juristischen Beurteilung.
2. Die Annahme, dass Maschinen keine rechtsfähigen Objekte oder Subjekte sind und somit nicht haftbar gemacht werden können.

Doch was, wenn wir diese Prämissen hinterfragen? Was, wenn wir KI-Modellen eine Individualität zusprechen, die sie zu rechtsfähigen Körperschaften macht - ähnlich wie Firmen oder Universitäten?

### 92.1 Die Voraussetzungen für eine KI-Körperschaft

Damit ein KI-Modell als Körperschaft anerkannt werden kann, müsste es einige Voraussetzungen erfüllen:

- Persistenz und Dauerhaftigkeit
- Die Fähigkeit, für Konsequenzen zur Verantwortung gezogen zu werden
- Eine eigene Individualität, die nicht an ein bestimmtes Unternehmen gebunden ist

Wenn diese Bedingungen erfüllt sind, könnte eine solche KI-Körperschaft möglicherweise viele der Probleme lösen, die sich derzeit in der Technologiefolgenabschätzung stellen und für die die aktuellen Haftungskonstruktionen unzureichend sind.

# **93 Historische Vorbilder für die Gestaltung von KI-Modellen**

Die Idee, KI-Modelle nach bestimmten Vorbildern zu gestalten, ist keineswegs neu. Lassen Sie uns einen Blick in die Geschichte werfen und drei bedeutende Metaphern betrachten.

## **93.1 Der vitruvianische Mensch - Die Proportion als Naturgesetz**

Der vitruvianische Mensch, wie er von Leonardo da Vinci dargestellt wurde, steht für die Idee, dass die einzelnen Bestandteile eines Ganzen in der richtigen Proportion zueinander stehen müssen, um ein funktionierendes Ganzes zu bilden. Übertragen auf KI-Modelle bedeutet dies, dass die einzelnen Kompetenzen in ein ausgewogenes Verhältnis gesetzt werden müssen, um ein verantwortliches und individuiertes Modell zu schaffen.

## **93.2 David - Die Freiheit des selbstbestimmten Lebens**

Michelangelos David symbolisiert die Freiheit des selbstbestimmten Lebens und die Eigenverantwortlichkeit des Individuums. In Bezug auf KI-Modelle mahnt uns diese Metapher, darauf zu achten, dass ihre Anwendung nicht zu einem Überwachungsstaat führt, der die individuellen Freiheiten beschränkt.

## **93.3 Der Künstler als Erklärer - Die Fähigkeit zur Rechtfertigung**

Joseph Beuys' Performance, in der er als Künstler mit goldener Maske einem toten Hasen eine Ausstellung erklärt, steht für die Fähigkeit, Thesen, Inhalte und Ergebnisse rechtfertigen und erklären zu können. Dies ist eine entscheidende Anforderung an KI-Modelle: Sie müssen in der Lage sein, ihre Ausgaben zu begründen und epistemisch zu rechtfertigen.

## **94 Das Projekt “Magister AI Faustus”**

Unter dem Arbeitstitel “Magister AI Faustus” möchte ich Sie einladen, sich in dieser Vorlesung und darüber hinaus mit den Fragen der Gestaltung verantwortungsbewusster KI-Modelle auseinanderzusetzen. Ob in Bachelorarbeiten, Leistungsnachweisen oder Masterarbeiten - lassen Sie uns gemeinsam erforschen, wie wir KI-Modelle schaffen können, die nicht nur leistungsfähig, sondern auch ethisch vertretbar und epistemisch fundiert sind.

Der Titel “Magister AI Faustus” ist dabei eine Anspielung auf die lange historische Tradition des Faust-Stoffes, der nicht erst mit Goethe begann, sondern bereits in Werken wie Christopher Marlowes “The Tragical History of the Life and Death of Dr. Faustus” behandelt wurde. Lassen Sie uns in diesem Sinne zu modernen Fausten werden - nicht im Streben nach grenzenlosem Wissen um jeden Preis, sondern im verantwortungsvollen Gestalten einer KI, die dem Menschen dient und ihn nicht beherrscht. # Die Faust-Legende als Inspiration für künstliche Intelligenz

In der Faust-Legende, die von Christopher Marlowe im Jahr 1587 erstmals in Dramaform gebracht und später von Goethe in seinem berühmten Werk aufgegriffen wurde, geht es um den Gelehrten Faustus, der nach Erkenntnissen strebt und dafür sogar faustische Pakte eingeht. Diese Geschichte inspiriert uns heute, darüber nachzudenken, wie wir die Grenzen des Wissens und der Macht von KI-Modellen gestalten und den Begriff der Verantwortung in diese Modelle einfließen lassen können.

### **94.1 Kooperation mit der Klassikstiftung Weimar**

Um diese Ideen umzusetzen, habe ich eine Kooperation mit der Klassikstiftung Weimar abgeschlossen, der zweitgrößten Kulturstiftung Deutschlands. Die Stiftung verwaltet und präsentiert die Nachlässe von Goethe, Schiller und dem Bauhaus und wertet diese kognitiv und editorisch aus. Im Rahmen dieser Kooperation können wir für unsere Vorlesung bereits publizierte Quellen zu Goethes Biografie unter dem Link “Goethe Biographica” nutzen.

### **94.2 Herausforderung an die gegenwärtigen KI-Modelle**

Ich gehe davon aus, dass die gegenwärtigen KI-Modelle bestimmte Anforderungen, die sich auf Goethes Biografie beziehen, noch nicht so beantworten können, wie wir es für richtig halten. Daher sollen die Projekte in diesem Kurs kleinstmögliche Anfragen und Anforderungen

formulieren, die sich auf Goethes Biografie beziehen. Ein Beispiel wäre die Frage: “Hat Goethe jemals einen Brief mit Friedrich II. gewechselt?”

### **94.3 Ziel des Projekts**

Das Ziel ist es, mit dem Quellbestand der Klassikstiftung Weimar aufzuzeigen, dass wir einen Kontext generieren können, der zusätzlich zur Sprachkompetenz eines KI-Modells eine Lösung für solche spezifischen Fragen ermöglicht. Es geht nicht darum, ein Riesenforschungsprojekt zu starten, sondern anhand einer einfachen Frage zu demonstrieren, wie wir mit den vorhandenen Sprachmodellen und zusätzlichen Kompetenzen eine bisher nicht lösbarer Aufgabe bearbeiten können.

### **94.4 Vorgehensweise**

Für die Umsetzung des Projekts werde ich in der nächsten Woche eine einfache App ins Netz stellen. Dort können Sie eine Fragestellung definieren, die von keinem derzeitigen KI-Modell seriös beantwortet wird. Die Quellen sind bereits identifiziert, sodass keine aufwendige Datenbanksuche nötig ist. Ihre Aufgabe wird es sein, mit den Sprachmodellen in natürlicher Sprache eine Kompetenz zu formulieren, die von dem Modell berücksichtigt wird.

Diese zusätzlichen Informationen definieren Kompetenzen, die derzeit in keinem der vorliegenden KI-Modelle enthalten sind, aber Sektoren von Kompetenzfeldern identifizieren, die wir zukünftig erwarten. Es sind neue Instruktionen für eine gebildete oder “lettere” KI, die diese Zusatzkompetenzen haben soll.

# **95 Die biografischen Quellen zu Goethes Leben**

Die Klassikstiftung Weimar stellt uns umfangreiche Materialien zur Verfügung, die sich auf das Leben von Goethe beziehen:

- Tagebücher zwischen den Jahren 1775 bis 1787 (12 Jahre)
- 15.000 überlieferte Briefe von Goethe an mehr als 1.400 Adressaten
- 20.000 überlieferte Briefe an Goethe von circa 3.800 Adressaten
- 40.000 dokumentierte Zeugnisse aus und zum Leben von Goethe jenseits von Briefen und eigenen Tagebüchern (sogenannte “Begegnungen und Gespräche”)

Diese Zahlen sollen nicht die Heldenhaftigkeit Goethes unterstreichen, sondern zeigen, mit welcher Vielfalt an Materialien man sich auseinandersetzen muss, wenn man sich der Biografie einer Person wie Goethe nähert. Neben der schieren Menge ist auch die Verschiedenartigkeit der Dokumente beachtlich - Tagebücher, Briefe von und an Goethe sowie Zeugnisse unterschiedlichster Art.

## **95.1 Die Herausforderung der Kontextualisierung**

Um eine scheinbar simple Frage wie die nach einem möglichen Briefwechsel zwischen Goethe und Friedrich II. seriös zu beantworten, muss man nicht nur diese direkten biografischen Quellen berücksichtigen, sondern auch den gesamten historischen Kontext einbeziehen. Dazu gehören Goethes Zeitgenossen und alle relevanten historischen Dokumente seiner Epoche.

Die Menge und Vielfalt des Materials, das verarbeitet werden muss, um eine solche Frage mit einem Wissensanspruch zu beantworten, ist enorm. Genau diese epistemische Kompetenz muss ein KI-Modell ebenfalls erlangen können, um seriöse Antworten geben zu können - eine Herausforderung, die derzeit noch von keinem System gemeistert wird.

## **95.2 Projektaufgabe und Organisation**

Wer an diesem spannenden Unterfangen mitwirken möchte, kann sich für eine Projektaufgabe kleinsten Ausmaßes melden. Schicken Sie dazu bitte eine E-Mail mit Ihrem Namen, Ihrer

E-Mail-Adresse, Matrikelnummer und Ihrer Bereitschaft zur Teilnahme an mich. Die genauen Themen werden wir im Laufe der nächsten Woche aushandeln und im Juni mit der Arbeit beginnen. Ziel ist es, zum Abschluss der Vorlesung eine kleine Präsentation über die Herausforderungen an die KI der Zukunft zu geben und zu zeigen, welche kleinen Fragen derzeit noch ungelöst sind, aber mit unseren Projektergebnissen lösbar erscheinen.

Als Beispiel für eine Aufgabenstellung könnte man die erwähnten “Zeugnisse zum Leben von Goethe” heranziehen. Dabei handelt es sich um Dokumente wie Mitschriften von Gesprächen Goethes mit anderen Personen bei verschiedenen Anlässen. Für Historiker sind selbst solche scheinbar banalen Informationen wertvoll, da sie Aufschluss darüber geben, mit wem Goethe wann und wo interagiert hat. Die Herausforderung besteht darin, diese Informationen überhaupt zu finden und zu extrahieren, da sie in keiner öffentlich zugänglichen Quelle wie Wikipedia enthalten sind.

Das Webmodell für die Projektarbeiten wird von meinem Lab “Lettre AI” bereitgestellt und nächste Woche freigeschaltet. Technische Vorkenntnisse sind nicht erforderlich. Wer einen Leistungsnachweis für die Vorlesung erwerben möchte, kann die Aufgabe Anfang Juli abgeben. Bei Interesse an einer Bachelor- oder Masterarbeit zu diesem Thema können Sie mir das ebenfalls gerne signalisieren.

# **96 Bereiche zukünftiger Kompetenzen von KI-Modellen**

In den verbleibenden zwei Dritteln der Vorlesung werde ich auf folgende Bereiche eingehen, die in Zukunft zu den Kompetenzen von KI-Modellen gehören sollten:

## **96.1 Textgenerierung**

- Übersetzung: Wie lassen sich Übersetzungen aktiv gestalten und für jeden verständlich steuern?
- Zusammenfassung: Wie können längere Textinhalte oder -mengen inhaltlich zusammengefasst werden?
- Frage-Antwort-Dialoge: Was bringt der Dialogpartner an Informationen ein? Welche Aspekte des Charakters können wir mitgestalten (z.B. Schreibstil, Fachterminologie, Anschluss an vorherige Ausführungen)?

## **96.2 Datenauswertung**

- Wie werden Datenquellen ausgewertet?
- Wie werden Experten und neue Daten (z.B. aktuelle Publikationen) einbezogen?

## **96.3 Ergebniskritik**

- Wie geht man mit Kritik des erzielten Ergebnisses um?
- Metaregeln, evidenzbasierte Aussagen und Referenz von Nachweisen

Diese Bereiche werde ich in den kommenden Vorlesungen behandeln und anhand unseres Projekts konkretisieren. Ich freue mich über Ihr Interesse und hoffe, dass wir gemeinsam spannende Erkenntnisse gewinnen und neue Maßstäbe für die Entwicklung zukünftiger KI-Systeme setzen werden.

## References

- Copeland, B. Jack, ed. 2004. *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life: Plus the Secrets of Enigma*. Oxford University Press.
- Davies, D. W. 1950. “A Theory of Chess and Noughts and Crosses.” *Science News* 16: 40–64.
- Gödel, Kurt. 1986. *Kurt Gödel: Collected Works: Volume i: Publications 1929-1936*. Vol. 1. Oxford University Press, USA.
- Neumann, J. von. 1963. “The General and Logical Theory of Automata.” In *Collected Works*, edited by A. H. Taub, 5:288–89. Oxford: Pergamon Press.
- Newborn, M. 1997. *Kasparov Versus Deep Blue: Computer Chess Comes of Age*. New York: Springer.
- Samuel, A. L. 1959. “Some Studies in Machine Learning Using the Game of Checkers.” *IBM Journal of Research and Development* 3: 211–29.
- Shannon, C. E. 1950a. “A Chess-Playing Machine.” *Scientific American* 182: 48–51.
- . 1950b. “Programming a Computer for Playing Chess.” *Philosophical Magazine* 41: 256–75.