

Auf den Spuren von Turing

Kurzbericht zum Turing-Test 2017, Christian Pallay

Bereits 1950 stellte der – oftmals als Vater der *künstlichen Intelligenz (KI)* deklarierte – Brite Alan Mathison Turing die Frage: “Können Maschinen denken?” Zur Beantwortung schlug er unter der Bezeichnung *imitation game* das folgende Experiment vor: Ein Juror führt je eine maschinenschriftliche Kommunikation – einen Chat – mit einem Menschen und einer Maschine. Sollte der Juror nach einer gewissen Konversationszeit Mensch und Maschine nicht unterscheiden können, wird der Maschine Intelligenz zugesprochen. Das heute in der Regel als *Turing-Test* bezeichnete Verfahren war einige Jahrzehnte zunächst nur Thema philosophischer Debatten.

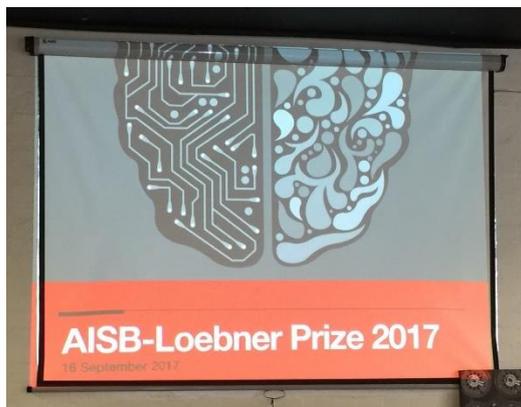
Seit 1991 findet jedoch im Rahmen des sog. *Loebner-Preises* jährlich eine empirische Realisierung statt. 2017 wurde dem Autor im Rahmen der Internationalisierungsmaßnahme der TUM die Gelegenheit gegeben, als Juror an der Veranstaltung teilzunehmen. Im Vorfeld wurden online über einen Auswahlprozess, in dem die 16 teilnehmenden Chatbots jeweils 20 identische Fragen beantworten mussten, vier Finalteilnehmer bestimmt. Ergänzt wurden diese um die gleiche Anzahl menschlicher Kandidaten. Letztlich wurden noch vier Juroren bestellt, darunter zwei Medienschaffende und zwei Forschende aus dem Bereich der KI, einer davon der Autor. Das Finale des Wettbewerbs wird seit einigen Jahren im Bletchley Park im Vereinigten Königreich ausgetragen. Heute ein sehenswertes Museum, wurde hier im zweiten Weltkrieg der Nachrichtenverkehr der deutschen Kriegsgegner entziffert. Einer der zentralen Kryptoanalytiker damals: Alan Mathison Turing.

Am dritten Septemberwochenende war es soweit: Über Terminals wurden vier Chat-Sitzungen von je 25 Minuten abgehalten. Jeder Juror konversierte pro Runde mit je einem Menschen und einer Maschine – ohne zu wissen wer sich hinter welchem Chatfenster befindet. Am Ende jeder Sitzung mussten die Juroren entscheiden, welchen ihrer beiden Gesprächspartner sie für den Menschen und welchen für die Maschine halten. Die Qualität der Interaktionen variierte beträchtlich: Während zwei der maschinellen Partizipanten mit massiven technischen Problemen zu kämpfen hatten und ihre Äußerungen ungewollt fortwährend wiederholten, waren die anderen beiden Chatbots in Teilen überzeugende Gesprächspartner. Letztlich wurden aber auch diese überführt, indem die Juroren sie mit mehrdeutigen Fragen oder komplexen Themen aus dem Zeitgeschehen überforderten. Es war daher auch kein überraschendes Ergebnis, dass keine Maschine die Illusion lange genug aufrechterhalten konnte, ein Mensch zu sein: 25 Minuten glaubhafter Dialog liegen noch nicht im Rahmen der Fähigkeiten der aktuellen Systeme natürlicher Sprachverarbeitung. Dies lässt sich auch privat leicht mit kommerziellen Sprachassistenten wie *Siri*, *Alexa* oder *Google Home* nachvollziehen. Über ein sekundäres Ranking wurden die maschinellen Beiträge schließlich noch untereinander bewertet. Am Ende konnte sich *Mitsuku* als der überzeugendste Chatbot durchsetzen und ging damit als relativer Gewinner hervor. Dafür durfte sich der Programmierer Steve Worswick über eine Medaille und USD 4000 freuen.

Auch 2017 zeigten sich einige der charakteristischen Schwächen des Turing-Tests. Um zwei zu nennen: Die Resultate hängen entscheidend von den Vorkenntnissen der Juroren ab und sind daher nicht objektiv. Zum Nachteil der Maschinen waren in diesem Jahr alle Juroren mit der Materie des Experiments vertraut. Problematisch ist auch die ausschließlich binäre Entscheidung zwischen Mensch und Maschine – repräsentativ für intelligent/nicht-intelligent – nach jeder Runde. Angebracht wäre

vielmehr ein gradueller Intelligenzbegriff, der die Realität widerspiegeln würde. Das sekundäre Ranking schafft hier nur bedingt Abhilfe.

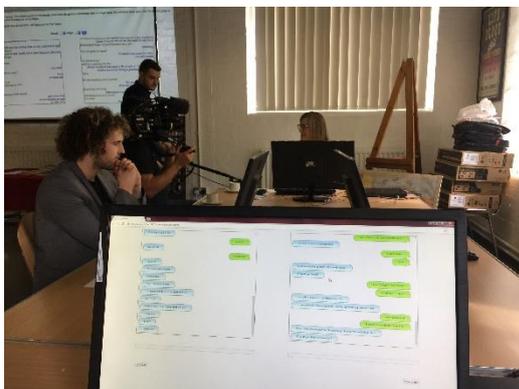
Im Rahmen seiner Dissertation entwickelt der Autor eine Systematik von Verfahren zur Messung von Maschinenintelligenz und verfolgt die Umsetzung einer objektiven, universellen und empirisch realisierbaren Methode, die Unzulänglichkeiten des Turing-Tests und anderer existierender Prozeduren beheben soll. Durch die Internationalisierungsmaßnahme konnten diesbezüglich insbesondere Einsichten in die praktische Umsetzung eines Verfahrens zur Evaluation von intelligenten Maschinen gewonnen werden. Der Dank dafür gilt der *AISB (The Society for the Study of Artificial Intelligence and Simulation for Behaviour)* für die Organisation und Einladung, allen Teilnehmern für deren Partizipation und besonders dem *Internationalization Grant der TUM-GS* für die Förderung.



Loebner Prize 2017



Bletchley Park, Hauptgebäude



Arbeitsplatz der Juroren



Siegermedaille

Website der AISB zum *Loebner-Preis* 2017:

<http://www.aisb.org.uk/events/loebner-prize>

Website aller Chatprotokolle – der Autor trat als *Judge 3* auf – des Finales:

<http://aisb-loebner-prize.org/static/webcast.html>

Website des Sieger-Chatbots Mitsuku inkl. der Möglichkeit mir einer älteren Version zu chatten:

<http://www.mitsuku.com/>