

Dekodierung der Kommunikation von nicht-menschlichen Spezies mit Hilfe von künstlicher Intelligenz

Aktuelle Themen der künstlichen Intelligenz
Sommersemester 2023
Teresa Mary Kutzner, Matr.-Nr. 45244

1 Abstract

Dieses wissenschaftliche Paper untersucht das aufstrebende Forschungsfeld der Dekodierung der Kommunikation von nicht-menschlicher Spezies unter Verwendung von künstlicher Intelligenz. Die Menschheit hat schon lange das Bedürfnis, die Kommunikation mit anderen Spezies zu verstehen. Dieses Interesse hat zu diversen Projekten geführt, darunter Project CETI und das Earth Species Project. Die Forschung zielt darauf ab, Kontexte und Muster in der nicht-menschlichen Kommunikation zu identifizieren, um sie in unsere menschliche Sprache zu übersetzen und so die Wahrnehmung und den Schutz unterschiedlicher Tierarten zu verbessern. Dieses Paper beleuchtet die Technologien und Herangehensweisen, die in diesem Bereich verwendet werden, sowie die aktuellen Forschungsergebnisse und ethischen Herausforderungen.

2 Einführung

Menschen zeichnen sich durch ihre ausgeprägte Neigung zur Kommunikation aus. Dieses Bedürfnis, Informationen auszutauschen, erstreckt sich nicht nur auf die Interaktion mit Artgenossen, sondern auch auf das Verlangen, die Kommunikation mit anderen Arten zu verstehen [1].

Dieses Interesse hat die Entstehung eines aufstrebenden Forschungsfeldes zur Folge gehabt: die Entschlüsselung der Kommunikation nicht-menschlicher Spezies, ein Bestreben, das sich auf die Enthüllung und Interpretation von Kommunikation abseits der menschlichen Sprache konzentriert. In diesem Kontext setzen Forschende vermehrt auf künstliche Intelligenz (KI), um diese Herausforderung zu bewältigen [2].

Die Frage, warum Wissenschaftler und Forscher ein so starkes Interesse an der Kommunikation und Übersetzung der Signale von Tieren hegen, bezieht sich weniger auf die Kommunikation mit Haustieren, sondern vielmehr auf die Interaktion mit anderen intelligenten

Spezies, die den Menschen in ihrer Intelligenz und Kommunikationsfähigkeit ähneln oder diese sogar übertreffen.

Ein bezeichnendes Beispiel für die Motivation hinter dieser Forschung ist die Untersuchung von Pottwalen. Diese Wal-Spezies zeichnet sich durch ihr beachtliches Gehirnvolumen aus und verfügt über eine eigene, komplexe Form der Kommunikation, die "Codas" und Klickgeräusche umfasst, was es den Tieren ermöglicht, individuelle Namen zu tragen und miteinander zu kommunizieren. Die Wale leben in sozialen Gemeinschaften, bestehend aus Clans und Familien, und pflegen Beziehungen innerhalb dieser Gruppen. Neben diesen Faktoren legt die Tatsache, dass diese Spezies über 20 Millionen Jahre alt ist, während die Menschheit im Vergleich dazu nur 300.000 Jahre existiert, nahe, dass sie eine eigene, einzigartige Kultur und Sprache entwickelt haben, die ihre Wahrnehmung der Umwelt stark beeinflusst [2].

Aufgrund dieser faszinierenden Besonderheiten bei vielen unterschiedlichen Spezies sind Wissenschaftler*innen besonders motiviert, die Kommunikation und damit das Wissen bestimmter Tierarten zu erforschen. Das übergeordnete Ziel dieser Forschung besteht darin, die Kontexte, Muster und Informationen in der Kommunikation nicht-menschlicher Spezies zu identifizieren und in unsere menschliche Sprache zu "übersetzen". Dies ermöglicht ein tieferes Verständnis der Umwelt und der Wahrnehmung dieser Spezies, was wiederum dazu beitragen kann, Lehren zu ziehen und Schutzmaßnahmen zum Erhalt diverser Spezies zu entwickeln [3].

3 Aktuelle Forschung

Das Feld der Erforschung der nicht-menschlichen Kommunikation befindet sich noch in seinen Anfängen und erfährt einen stetigen Zuwachs an Forschungsteams, die sich diesem Thema widmen. Besonders herausragende Fortschritte werden durch zwei wegweisende Projekte erzielt: das Project CETI und das Earth Species Project.

Project CETI wurde im Jahr 2020 als eine gemeinnützige Organisation auf der karibischen Insel Dominica in Südamerika gegründet. Die Forscher*innen dieses Projekts haben bereits seit geraumer Zeit intensive Untersuchungen an den dort heimischen Pottwal-Familien durchgeführt, wodurch sie über bedeutende Datensätze verfügen, die für die Kommunikationsforschung von großer Relevanz sind. CETI erhält breite Unterstützung von einer Vielzahl verschiedener Partner, darunter namhafte Institutionen wie die Harvard- und Oxford-Universitäten sowie Organisationen wie National Geographic und Google Research. Das vorrangige Ziel von CETI besteht darin, die Kommunikation der Pottwale weiterhin intensiv zu erforschen, um Daten zu sammeln und mithilfe von künstlicher Intelligenz die Geheimnisse ihrer Kommunikation zu entschlüsseln [4].

Das Earth Species Project wurde bereits im Jahr 2013 in Kalifornien, USA, gegründet und verfolgt ebenfalls das Ziel, die Kommunikation nicht-menschlicher Spezies zu entschlüsseln. Im Gegensatz zu Project CETI konzentriert sich das Earth Species Project nicht ausschließlich auf Wale, sondern widmet sich auch intensiv anderen Tierarten wie afrikanischen

Elefanten, Fledermäusen und Krähen. Das Forschungsteam des Earth Species Project erhält Unterstützung von denselben und weiteren Partnern wie CETI, verfügt jedoch über zusätzliche Datenquellen, um die große Vielfalt der zu erforschenden Spezies abzudecken [5].

4 Technologie

4.1 Anforderungen

Um die Entschlüsselung nicht-menschlicher Kommunikation erfolgreich durchzuführen, ist es von entscheidender Bedeutung, die Rahmenbedingungen und Herausforderungen innerhalb dieses Forschungsbereichs zu betrachten.

Die Kommunikation nicht-menschlicher Spezies erstreckt sich über vielfältige Kanäle. Dabei spielen nicht nur Audiosignale eine Rolle, sondern auch Körpersprache und der Kontext, in dem diese Signale auftreten. Diese vielseitigen Kommunikationskanäle sind entscheidend für die Bedeutung und Interpretation einer Äußerung und stellen eine der ersten Herausforderungen für das Forschungsfeld dar.

Dabei ist außerdem zu beachten, dass sich Syntax und Kontext nicht-menschlicher Kommunikation erheblich von menschlicher Sprache unterscheiden. Dies führt zu Unsicherheiten und Herausforderungen bei der Identifizierung und Interpretation von Mustern und Bedeutungen der Aussagen [6].

Des Weiteren gestaltet sich die Sammlung von Daten in natürlichen Lebensräumen nicht-menschlicher Spezies äußerst anspruchsvoll. Die Umweltbedingungen und die scheuen Verhaltensweisen vieler Tiere erschweren die Datenerfassung

erheblich. Darüber hinaus besteht die Gefahr, dass die gesammelten Daten beschädigt oder durch andere Einflüsse verfälscht werden, was die Analyse und Interpretation beeinträchtigen kann. Es ist daher notwendig, die Daten unter Berücksichtigung des Kontexts umfassend zu sammeln und dafür die nötigen Technologien zur Verfügung zu haben [7].

Eine weitere Anforderung an das Forschungsfeld entsteht durch ethische Bedenken von Expert*innen hinsichtlich der Kommunikation mit anderen Spezies. Diese Bedenken beziehen sich auf den menschlichen Einfluss auf die Lebensräume und Kulturen nicht-menschlicher Spezies. Im weiteren Verlauf dieses Berichts werden diese ethischen Aspekte näher erörtert und diskutiert [3].

4.2 Herangehensweise

Das Earth Science Project hat in Abbildung 1 eine effektive Vorgehensweise zur Dekodierung nicht-menschlicher Kommunikation veranschaulicht [8]. Diese Methode umfasst mehrere Schritte:

1. Datensammlung in der Wildnis

Der erste Schwerpunkt liegt auf der umfassenden Datensammlung aus natürlichen Lebensräumen der zu untersuchenden Spezies.

2. Training eines Foundation Modells

Anschließend erfolgt das Training eines sogenannten Foundation Modells mithilfe der gesammelten Daten. Ein Foundation Modell ist ein Machine Learning Modell, das auf umfangreichen Datensätzen trainiert wurde und anschließend an verschiedene Aufgaben angepasst oder

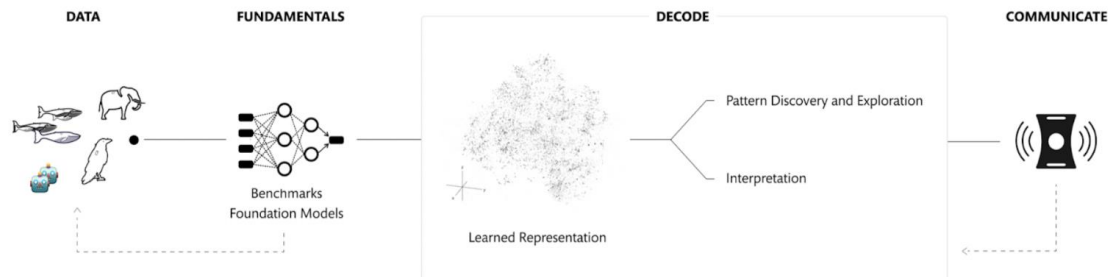


Abbildung 1: Roadmap des Forschungsprojekts Earth Science Project. Quelle: Earth Science Project, [8]

feinabgestimmt werden kann. Beispiele für aktuelle Foundation Modelle sind BERT, GPT-4 und CLIP [12].

3. Erstellung von Benchmarks

Während des Trainingsprozesses des Modells werden Benchmarks entwickelt, um die Leistung des Modells anhand verschiedener Leistungsmetriken zu bewerten. Dies ermöglicht es, das Modell im Kontext des Forschungsfelds mit anderen Modellen zu vergleichen.

4. Iterative Anpassung des Foundation Modells

Das Foundation Modell wird wiederholt an die gesammelten Daten angepasst, bis es eine gute Leistung gemäß den Benchmarks erreicht.

5. Erlernen des Wortschatzes der Spezies

Das Modell erlernt den spezifischen Wortschatz der zu untersuchenden Spezies aus den gesammelten Daten und speichert diesen als erlernte Repräsentation.

6. Suche und Vergleich von Mustern

In dieser erlernten Repräsentation werden Muster gesucht und mit menschlicher Sprache verglichen, um Übereinstimmungen zu finden und Kontexte zu erfassen.

7. Kommunikation mit der Spezies

Nach erfolgreicher Dekodierung der gesammelten Daten soll dann die Kommunikation mit der untersuchten Spezies folgen. Das erhaltene Feedback wird verwendet, um das Modell und seine Repräsentation weiter zu verfeinern und zu trainieren.

Im folgenden Abschnitt werden die Datensammlung und die Grundlagen des Foundation Modells näher erläutert, um zu verdeutlichen, wie es möglich ist, aus Daten eine Repräsentation einer Sprache zu erlernen.

4.2.1 Sammeln der Daten

Die Datensammlung in der Tierwelt erfordert die Berücksichtigung zweier wesentlicher Faktoren: Verfügbarkeit und Vielseitigkeit [3].

Um erfolgreich Daten zu sammeln, müssen diese in ausreichender Menge verfügbar sein. Dies bedeutet, dass eine kontinuierliche Datenerfassung über einen längeren Zeitraum notwendig ist, um relevante Einblicke in die Kommunikation nicht-menschlicher Spezies zu gewinnen [7].

Damit die Verfügbarkeit von Daten sichergestellt werden kann, werden unterschiedliche

Datenerfassungstechnologien eingesetzt. Dazu gehören an Tieren befestigte Logger, Drohnen, Schwimmroboter und Unterwasser-Mikrofone, um eine breite Palette von Datenquellen zu erschließen [4].

Die Verwendung verschiedener Datenerfassungstechnologien hat den zusätzlichen Vorteil, die Vielseitigkeit der gesammelten Daten zu erhöhen. Das bedeutet, dass Informationen aus verschiedenen Perspektiven und in unterschiedlichen Formaten erfasst werden können. Dieser vielfältige Datenpool ist entscheidend für die Entwicklung und das Training von Modellen zur Dekodierung der nicht-menschlichen Kommunikation [4].

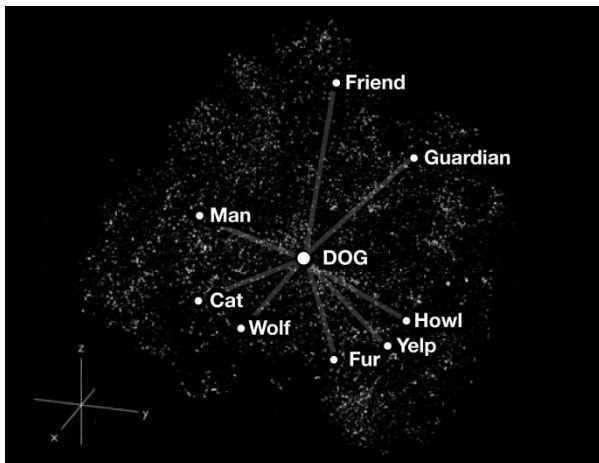


Abbildung 2: Darstellung der Wörter der englischen Sprache in einem vereinfachten 3-dimensionalen Raum. Quelle: Google, [3]

Durch die Kombination von verschiedenen Technologien und die Sicherstellung einer umfangreichen Datenvielfalt wird gewährleistet, dass ausreichend Daten zur Verfügung stehen, um ein leistungsfähiges

Modell zur Kommunikationsdekodierung zu entwickeln und zu trainieren.

4.2.2 Grundlagen des Foundation Models

Um ein KI-Modell in die Lage zu versetzen, die Sprache einer beliebigen Spezies zu erlernen, ist es zunächst wichtig, zu verstehen, wie ein Machine Learning Modell solche Fähigkeiten aus Daten erwerben kann. Aza Raskin, Mitbegründer des Earth Science Projects, beschreibt diesen Lernprozess anschaulich mit dem Satz: "AI turns semantic relationships into geographic relationships." [3].

Diese Aussage bedeutet, dass ein Machine Learning Modell die einzelnen Wörter oder Laute einer Sprache in einem multidimensionalen Raum abbildet. Ein Beispiel dafür ist in Abbildung 2 zu sehen, wo die englische Sprache vereinfacht in drei Dimensionen dargestellt ist.

In diesem Raum sind die einzelnen Wörter so positioniert, dass ihre räumliche Nähe zu anderen Wörtern deren Kontext und Beziehungen beschreibt [3].

Zum Beispiel wird das Wort "Hund" in der Nähe von "Katze" oder "Wolf" positioniert, aber es zeigt auch Verbindungen zu Wörtern wie "Freund" oder "Beschützer", die weiter entfernt sind. Diese geografische Darstellung ermöglicht es dem Modell, die Repräsentation einer menschlichen oder nicht-menschlichen Spezies abzubilden und durch Interpretation zu erlernen.

Abbildung 3 illustriert ein bemerkenswertes Phänomen, das während der Arbeit mit dieser Repräsentation aufgefallen ist. Verschiedene menschliche Sprachen, wie Spanisch, Englisch und Japanisch, weisen

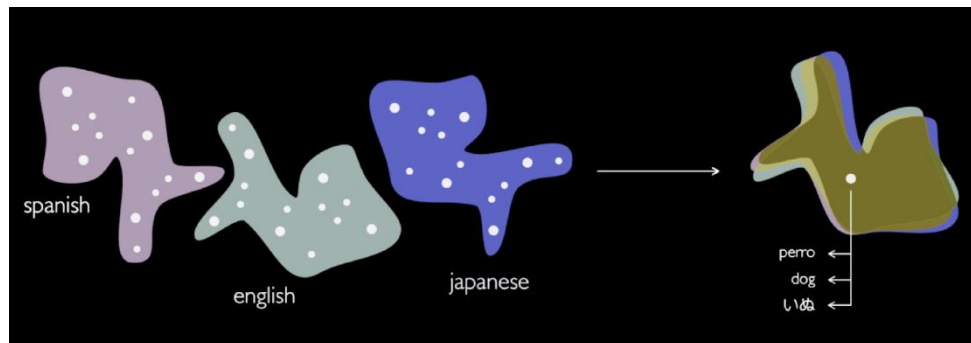


Abbildung 3: Darstellung der drei Sprachen Spanisch, Englisch und Japanisch als visuelle Repräsentationen mit derselben Form. Quelle: Google, [3]

eine ähnliche geografische Form auf, was bedeutet, dass einzelne Wörter an ähnlichen Stellen positioniert sind. Dies hat zur Folge, dass Sprachmodelle wie zum Beispiel GPT schnell die unterschiedlichen Kontexte von Wörtern in verschiedenen Sprachen erlernen können, wodurch unter anderem auch das Übersetzen von Sprachen mit Hilfe von KI ermöglicht werden kann [3].

Für die Forschung im Bereich nicht-menschlicher Spezies ist dieser Ansatz von entscheidender Bedeutung. Die Forschenden streben an, mithilfe des Foundation Modells eine vergleichbare Repräsentation für die Sprache von verschiedenen Spezies zu erstellen, um diese dann mit der menschlichen Sprache zu vergleichen. Hierbei ist von Interesse, welche Bereiche dieser "Wortwolken" übereinstimmen und somit direkt übersetzbar sind. Gleichzeitig sind die nicht übereinstimmenden Bereiche ebenfalls von bedeutendem Interesse, da diese auf bisher unbekanntes Wissen oder unbekannte Kontexte hinweisen könnten [3].

5 Aktuelle Forschungsergebnisse

Es gibt bereits aktuelle Forschungsergebnisse, welche von den Forschungsteams veröffentlicht wurden, welche im Weitergang erläutert werden.

Das Earth Science Project veröffentlichte im Oktober 2022 die Publikation "BEANS - The Benchmark of Animal Sounds", welche aktuelle Benchmarks und öffentliche Datensätze für die Arbeit mit bioakustischen Daten vorstellt [4]. Der Schwerpunkt liegt auf der Klassifikation und Erkennung von bioakustischen Daten, um Machine Learning Modelle anhand dieser Benchmarks besser vergleichbar und bewertbar zu machen. Dieser erste Schritt ist von großer Bedeutung, um in Zukunft verschiedene Foundation Modelle zu entwickeln und diese miteinander vergleichen zu können [9].

Abbildung 4 markiert einen bedeutsamen Meilenstein für das Earth Science Project: eine vorläufige geografische Darstellung verschiedener Laute der Walart Beluga. Wie im Abschnitt 4.2.2 erläutert, sind solche Repräsentationen entscheidend für den Erfolg der Dekodierung nicht-menschlicher Sprache. Der hier gezeigte Wortschatz stellt lediglich den Anfang dar

und wird in den kommenden Jahren kontinuierlich erweitert [10].



Abbildung 4: Darstellung des Sprachschatzes der Walspezies Beluga des Earth Science Projects.
Quelle: Vimeo, [10]

Zusätzlich zu diesen Fortschritten sind im Zuge der Forschung zur Dekodierung nicht-menschlicher Sprache weitere Herausforderungen aufgetreten, die auch für andere Forschungsbereiche im Bereich der Sprache relevant sind.

Ein Beispiel ist das sogenannte "Cocktail Party Problem", bei dem die Forschenden des Earth Science Projects daran arbeiten, Möglichkeiten zu finden, wie KI verwendet werden kann, um Audiospuren zu unterscheiden, in denen mehrere Stimmen gleichzeitig zu hören sind [11].

Des Weiteren wird die Forschung an Loggern, Sensoren und Mikrofonen durch dieses Forschungsfeld und seine Erkenntnisse kontinuierlich vorangetrieben [7].

Ebenfalls wächst die Forschung in Zusammenarbeit mit den untersuchten Tierarten, wobei versucht wird, mithilfe der erlernten nicht-menschlichen Sprachen mit Tieren zu kommunizieren. Dies wirft jedoch auch ethische Bedenken

auf, die im weiteren Verlauf genauer betrachtet werden [3].

6 Ethische Bedenken

Im Abschnitt 4.1 wurden bereits einige ethische Bedenken von Experten in Bezug auf dieses Forschungsfeld erwähnt. Diese Bedenken betreffen vor allem den Eingriff des Menschen in die Lebensräume und Kulturen anderer Spezies. Dieser Eingriff wird nicht nur durch das Sammeln von Daten und das Eindringen in die Lebensräume verursacht, sondern auch dann, wenn der Mensch versucht, mit anderen Spezies zu kommunizieren. Da die Kulturen einiger Tierarten noch unbekannt sind, ist unklar, inwieweit die Kommunikation diese Kulturen stört, und ihnen schadet [3].

Ein weiterer wichtiger ethischer Aspekt betrifft den verantwortungsvollen Umgang mit der potenziellen Kommunikation mit anderen Spezies. Es ist von großer Bedeutung, dass die Kommunikation ausschließlich zu Forschungszwecken erfolgt und nicht dazu genutzt wird, andere Interessen, wie beispielsweise wirtschaftliche oder industrielle, zu fördern. Es ist wichtig sicherzustellen, dass die Kommunikation nicht dazu missbraucht wird, andere Spezies auszubeuten oder zu schädigen [3].

7 Fazit

Das Forschungsfeld zur Dekodierung nicht-menschlicher Kommunikation mithilfe künstlicher Intelligenz ist bisher eher klein, verspricht jedoch vielversprechende Erkenntnisse über verschiedene Tierarten.

Diese Erkenntnisse können dazu beitragen, diese Arten besser zu schützen und bieten zudem neue Ansatzpunkte und Herausforderungen für die Weiterentwicklung von KI-Systemen im Bereich der Sprachverarbeitung. Dies zeigt, dass das Forschungsfeld bereits jetzt einen erheblichen Einfluss auf verschiedene andere Forschungsbereiche hat und daher von großer Bedeutung ist.

Durch die Erforschung unterschiedlicher nicht-menschlicher Spezies werden kontinuierlich neue Erkenntnisse gewonnen, die nach Ansicht der Forschenden von CETI und dem Earth Science Project in den kommenden Jahren zu erfolgreichen Fortschritten bei der Dekodierung nicht-menschlicher Sprachen führen könnten.

Das erklärte Ziel ist es, innerhalb der nächsten drei Jahre bereits in der Lage zu sein, mithilfe von KI mit anderen Spezies zu kommunizieren [3].

Quellenverzeichnis

- [1] Kulick, Don, Human–Animal Communication (October 2017). Annual Review of Anthropology, Vol. 46, pp. 357-378, 2017, Available at SSRN: <https://ssrn.com/abstract=3058760> or <http://dx.doi.org/10.1146/annurev-anthro-102116-041723>
- [2] Welch, C. (2021, April 21). Groundbreaking effort launched to decode whale language. National Geographic. <https://www.nationalgeographic.co.uk/animals/2021/04/groundbreaking-effort-launched-to-decode-whale-language>
- [3] Endler, M. (2023, April 21). Can generative AI lead people to understand animals? | google cloud blog. Google Cloud. <https://cloud.google.com/blog/transform/can-generative-ai-help-humans-understand-animals-earth-species-project-conservation?hl=en>
- [4] Project CETI – Website. Project CETI. <https://www.projectceti.org/>
- [5] Earth Species Project – Website. Earth Species Project. <https://www.earthspecies.org/>
- [6] Andreas J, Beguš G, Bronstein MM, Diamant R, Delaney D, Gero S, Goldwasser S, Gruber DF, de Haas S, Malkin P, Pavlov N, Payne R, Petri G, Rus D, Sharma P, Tchernov D, Tønnesen P, Torralba A, Vogt D, Wood RJ. Toward understanding the communication in sperm whales. *iScience*. 2022 May 13;25(6):104393. doi: 10.1016/j.isci.2022.104393. PMID: 35663036; PMCID: PMC9160774.
- [7] Self-supervised Ethogram Discovery. Earth Species Project. (2022, September 22). <https://www.earthspecies.org/blog/self-supervised-ethogram-discovery>
- [8] Our roadmap to decode. Earth Science Project. <https://www.earthspecies.org/what-we-do/roadmap>
- [9] M. Hagiwara, B. Hoffman, J. -Y. Liu, M. Cusimano, F. Effenberger and K. Zacarian, "BEANS: The Benchmark of Animal Sounds," ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1-5, doi: 10.1109/ICASSP49357.2023.10096686.
- [10] Beluga scatter plot. (2023). Vimeo. Earth Science Project. Retrieved September 13, 2023, from <https://vimeo.com/832004421>.

[11] Bee, Mark & Michey, Christophe. (2008). The Cocktail Party Problem: What Is It? How Can It Be Solved? And Why Should Animal Behaviorists Study It?. *Journal of comparative psychology* (Washington, D.C. : 1983). 122. 235-51. 10.1037/0735-7036.122.3.235.

[12] Bommasani, Rishi & Hudson, Drew & Adeli, Ehsan & Altman, Russ & Arora, Simran & Arx, Sydney & Bernstein, Michael & Bohg, Jeannette & Bosselut, Antoine & Brunskill, Emma & Brynjolfsson, Erik & Buch, Shyamal & Card, Dallas & Castellon, Rodrigo & Chatterji, Niladri & Chen, Annie & Creel, Kathleen & Davis, Jared & Demszky, Dora & Liang, Percy. (2021). On the Opportunities and Risks of Foundation Models.