第７講　未来への飛翔－生成ＡＩ、自動運転、ＡＩ倫理が紡ぐ社会の未来

【学習到達目標】

・日本のＡＩ戦略は教育から始まることを説明できる。
・教育に利用される生成ＡＩや自動運転について，事例を挙げて説明できる。
・国際的なＡＩのルール作りとＡＩ倫理について考えることができる。

１．日本のＡＩ戦略

政府見解として公開されたのが、２０１９年４月で「我が国のＡＩ原則/ＡＩ戦略について」（平井卓也・内閣府特命担当大臣）です。ＨｉｒＡＩ　Ｐｉｔｃｈにおけるａｉ関連テーマの例として「人材育成」、「社会実装」、「データ戦略」などがあり、ＡＩ戦略に関連する幅広いテーマについて産官学の関係者が意見交換いたしました。
具体的には、「ＡＩの中小企業向けサービス」、「ＡＩ検定試験」、「データ流通市場」や「衛星写真を組み合わせた都市の３Ｄマップ作成技術」がＡＩ関連テーマの例です。今日でいうデジタルツインというようなものが当時出てきたわけです。

・「ＡＩ戦略」の概要

２０２２年のＡＩ戦略では、「人間尊重」、「多様性」と「持続可能」の3つの理念と、ソサイアティ５．０を実現しＳＤＧｓに貢献するということを目標にしています。この３つの理念の実装念頭において、５つの戦略目標、「人材」、「産業競争力」、「技術体系」及び「国際」、それに加えて５つ目として、戦略目標０ということで「差し迫った危機への対処」というのを設定するようになりました。教育と関係するのは、「未来への基盤作り」の中の

「教育改革」です。
「産業・社会の基盤作り」項目では、「社会実装」が非常に重要されています。ＡＩ戦略２０２２での社会実装というのは、「他国の先進的な事例」との比較に基づいて、新たな目標を設定して推進するということです。

・人間中心のＡＩ社会原則の概要

ＡＩ戦略２０１９で言う「人間中心のＡＩ社会原則の概要」では、「人間中心の社会」を中心に周りに「人間の尊厳」、「多様性」と「持続可能性」というものがあります。
「人間中心の原則」というのは、ＡＩは人間の能力を拡張するが、ＡＩ利用に関わる最終判断は人が行うものです。教育と関係するのは「教育・リテラシーの原則」ですが、リテラシーを育む教育環境を全ての人々に平等に提供するとなっております。その他にも、「公正競争確保の原則」、「公平性、説明責任及び透明性の原則」、「イノベーションの原則」、「セキュリティ確保の原則」や「プライバシー確保の原則」があります。

・ＡＩ戦略の進捗状況と新たな戦略

ＡＩ戦略のこれまでの進捗状況と新たな戦略の必要性を示したものです。
ＡＩ戦略の主な成果の第１番目は、「教育改革」です。教育改革の１つ目は、「数理、データサイエンス、ＡＩ教育プログラム認定制度」が開始されたことです。これはリテラシーレベル（読み書きレベル）ですが、２０２１年までに７８件を認定するという成果を挙げています。
教育改革の２つ目は、皆さんも関わっておられる小学校、中学校における「ＧＩＧＡスクール構想」で、１人１台端末を前倒しで実現したということです。
他にもあの「人工知能研究開発ネットワーク」を設立してこととか、スマート農業実証プロジェクトとかが成果として現れてきました。
進捗としては、１７１件中１５４件、９０%が計画通りに進捗しています。

国外においては、ＡＩ技術は国家安全保障、民主主義保全などの社会の根本機能維持の必須技術となっています。

例えば、アメリカでは国家安全保障の観点から、ＡＩ予算（非国防）ということで年間３２０億ドル、プロ野球の大谷選手の契約金１０００億円超の約５０倍を、国家予算として軍事でない面だけでこれだけ投入しています。

一方、中国では、軍事に使うということで５か年計画を発表しています。他方、ＥＵはＡＩ利用に関する包括規制案を公表しています。

日本では、これまでの「社会・経済システムの変革」に加え、地震とか水害とか火山噴火という「大規模災害」への備えや、コビッド１９のような「パンデミックへの備え」に対応するような新たなＡＩ戦略が要求されます。

・国家規模の危機への対処

「国家規模の危機への対処」ということで、日本の場合は少子化ということで人口減少に伴う、我が国の体力の低下とか、デジタル化の遅れがあります。
これまでの閉塞を破る起爆剤として、ＡＩを大きく活用すべきというように捉えられているわけです。それには、２つの流れがあって、１つは最大速度でのデジタル化・ＡＩ化ということです。基幹インフラのＡＩ化前提でデジタルツイン、特に災害の場合に実際の街で水害とか火山の噴火できないので、デジタルツインの上でシミュレーションするというものです。
例えば、渋谷の街をそのまま再現して、デジタルで作った世界で、例えば人口がどう動くかとか？を見るわけです。
それからもう１つの流れは、強靭な社会システムへの転換ということで、グローバルネットワークの強化ということがあります。

・地球規模の危機への対処/強靭な基盤作り

規模の危機への対処」と「強靭な基盤作り」です。強靭な基盤作りということで、「責任あるＡＩの概念」を構築する「説明可能なＡＩ」等の技術は、情報基盤の信頼性を担保します。そして、我が国は高品種と安心安全という競争上の利点を生み出し、「責任あるＡＩ」に向けた取り組み行っていくことになります。
地球規模の危機への対処の方は、ＳＤＧｓで言うサステイナビリティ分野でのＡＩの応用を考えていくというわけです。

・ＡＩに関する「思い込み」

ＡＩの社会実装の推進に臨む姿勢には、３つの思い込みを捨てる必要があります。
１つ目は「ＡＩは人の仕事を代替する？」という思い込みです。ＡＩが我々の仕事を奪うのではと言うのは、思い込みであって、実はＡＩは人と協調します。人がＡＩと協調することで労力を最小化して利益を最大化することができます。
２０３０年雇用の大崩壊があり、ＡＩが人の仕事を奪うと言われています。

澤井は、日本の少子化を解決するため、むしろＡＩをどんどん使い込んで、生産技術では日本が世界で一番優れているとなったなら、日本の国際競争力が維持できると考えています。そんな訳で、人はＡＩと協調していく方が良いのではないかと思います。

２つ目の思い込みは「技術者だけがＡＩを深く理解できる？」です。これも思い込みです。最近のチャットＧＰＴとかＢＡＲＤとかＢＩＮＧ　ＡＩ等を見ると、「結構だれでも使える」感じです。ビジネスケースからＡＩは理解できるということです。自らＡＩを構築しなくても、既存のＡＩを利用し、他の部分で差別化していくことも一つの有効な手段です。

３つ目は、「データが全て？」です。今のＡＩはデータがなければ何も出来ないという思い込みがあります。実はそうではなく、使いながらどんどん増やしていけば良いというのが「ループの形成」です。ＡＩによるサービスの提供を通じてデータを取得するというようなループを形成し、サービスの提供とデータの収集を同時に行うことが重要です。

・ＡＩデータの課題

「データ」について、ＡＩ戦略２０１９に出た話ですが、ノイドや偏ったデータによってはＡＩが信頼できる結果を出すことができない可能性があります。

例えば採用試験を、ＡＩ採用システムがこれまでずっと過去の履歴では男性を取ることが多いとした場合、学習データのバイアスによって、女性を不適切な判断で不採用にするのではないか心配です。

また、少しのノイズでＡＩシステムが誤認識してしまいます。例えば最高時速５０と書いてある標識が泥で汚れてしまうと、速度制限５０以上出しても良いと読み間違えてしまうというような誤動作が発生します。

その意味でＡＩ製品のサービスの信頼性を担保する仕組みが必要です。その１つはの国際標準化で、いま１つは第三者評価プロセス、ガイドラインの整備です。

ＡＩ戦略２０２２では社会実装に向けた４つの取り組みがあります。

ディープラーニングを重要分野として位置づけ、企業による実装を念頭に置き、「ＡＩの信頼性の向上」、「人材確保の等の環境整備」、「ＡＩ利活用えるデータの充実」と「政府におけるＡＩ利活用の推進」に取り組みます。そして、物理、科学、機械など日本が強みを有する分野とＡＩの融合により、競争力の高い製品やサービスを生み出すことが今後大事になってきます。

・日本のＡＩ戦略のまとめ

日本のＡＩ戦略は３つにまとめることができます。１つは「「人間中心の社会原則の必要性」です。２つ目は、「ＡＩ社会原則を踏まえたＡＩ社会実装」で人間中心という意味を真に理解した上で実現すべき未来社会に向かうためのＡＩ社会実装を推進します。３つ目は、「我が国としての方向性」です。
人間中心に基づく健全なＡＩ社会実装で世界に先行し、広島サミットの宣言をおこなっていることです。２０２３年末ＡＩ戦略会議にて生成ＡＩに関する暫定的な論理を整理しています。これがまとまると次のＡＩ戦略になるというわけです。

２　生成ＡＩと教育

・ＡＩの１丁目１番は教育

ＡＩ戦略２０１９では、「ＡＩの１丁目１番は教育」ですと言っています。
登壇者は、座長の安西祐一郎・慶応義塾大学元学長、北野ソニー・コンピュータサイエンス研究所社長と内閣府の専門官です。
ＡＩ戦略２０１９での教育改革に関しする主な取り組みです。狙いは、デジタル社会の「読み・書き・そろばん」である「数理・データサイエンス・ＡＩ」の基礎などの必要な力を全国民が育み、あらゆる分野で活躍できるようにすることです。

・教育改革に関する主な取り組み

ＧＩＧＡスクール構想の前倒しで生徒１人１台端末もＩＣＴ環境これは実現されたので、今後小中学校では４校に１人以上、高校では１校に１人以上の多様なＩＣＴ人材を登用することになります。
「学習内容の強化」としては、高校におけるＡＩの基礎となる実習授業の充実で、大学でのＭＯＯＣ等を活用した標準カリキュラムの開発と展開です。２０２５年の育成目標は、小中学生全員と高校卒業生全員ということで毎年１００万人にしていこう、大学・高専卒業者は全員ということで毎年５０万人、合わせて１５０万です。
「応用基礎」は、高校の一部と大学の５０％で２０２５年には２５万人育成していこうというわけです。「認定制度・資格の活用」では、「大学等の優れた教育プログラムを政府が認定します。

「制度構築」し、全学生が受けれる講座を作ることです。この講座は著者だけじゃなくて何人かの先生方で一緒にすることです。更に「ＡＩ応用力の習得」ということで、ＡＩと専門分野、例えば文化系の文化創造であれば、ＡＩと文化創造とのダブルメジャーになるような人の育成を促進していくことになっています。「先鋭的な人材を発掘・伸ばす環境整備」ということでエキスパートを２０２５年には２０００人を育成する状況にして行きたいということです。

・生成ＡＩ／教育

２０２２年１１月にオープンＡＩが、ご存知のようにＣｈａｔＧＰＴを発表いたしました。爆発的な広がりを持って、２ヶ月で、１億人に達しました。ＣｈａｔＧＰＴは、事前学習した大規模なニューラルネットワーク（大規模言語モデル）を対話形式で操作する方法を採用し、「生成ＡＩ」として世界的に話題を呼びました。

生成ＡＩとは、全く新しいオリジナルのアウトプットを生み出すＡＩのことで、具体的には新しいデジタルの画像／動画、音声/音楽、文章／コード等を生成するＡＩ、もしくはこれらを組み合わせて生成するＡＩのことを指す、と言われています。
生成ＡＩは、非常に高性能で、コンピュータが我々の住む世界のことを良く理解し、人間と自然な対話（チャット）ができるような新しい能力を獲得しているかのように振る舞います。
生成ＡＩは、数億パラメータ以上の膨大な知識量を有し、答えを瞬時に提示することを得意としており、正解を提示する能力では人間を超えています。生成ＡＩは使い方次第で人間の創造的な作業を支援します。

・未来への基盤作り

教育改革に関する主な取り組みと、研究開発における主な取り組みです。「初等中等教育における環境整備」は、「１人１台端末」で実現されました。
リテラシーとして高等学校でのＡＩの基礎となる実習授業の充実と、大学等の優れた教育プログラムを政府が認定する制度で、毎年５０万人育成します。
応用基礎で毎年に２５万人、エキスパートと言われる人材は毎年２０００人育成していく計画です。
研究開発における主な取り組みでは、「人工知能研究開発ネットワークを理研、産総研とＮＩＣＴを中核にし、大学、海外の研究機関や国の研究機関とつぐネットワークを充実していきましょう」という計画です。

・生成ＡＩに関する暫定的な論点整理

以下の論点整理は、最近の技術の急激な変化や広島ＡＩプロセスを踏まえて、ＡＩ戦略会議構成員がＡＩ関連の論点を整理したものです。

〇生成ＡＩの可能性

生成ＡＩの登場は、内燃機関の発明・ＩＴ革命と同じく、幅広く生活の質を向上させる「歴史の転機」となる可能性があります。生産性の向上・情報アクセスの改善など、諸課題の解消も期待されています。

〇生成ＡＩと日本の親和性

我が国は、①研究・技術水準の高さ、②ロボット・ＡＩへの肯定的イメージ、③労働人口急減、④デジタル化への高いニーズ、⑤きめこまやかさ・創造性など、生成ＡＩとの親和性が高く、大きなチャンスです。アトムとかドラえもんとかロボットＡＩへの肯定的イメージが強くて、そこへ持ってきて、少子化による労働人口の急減があり、デジタル化の必要性が出てきています。きめ細やかさ創造性など、日本人の特性ともあって、生成ＡＩとの信和性が大きく、大きなチャンスになる可能性があるというわけです。

〇いま戦略を検討することの重要性

我が国に、ＡＩの勃興とともに再び成長の機運が見えており、諸外国の後塵を拝さないよう、今こそ大胆な戦略が必要です。政府は、人々がＡＩがもたらす社会変化に対して安心感を持ち、各プレイヤー、つまり各産業界の人たちが予見可能性を持てるようリスクに対応すべきという訳です。また、企業・研究者が存分に活動できるインフラ整備を行うべきです。例えばスパコン、ＡＩ用のスパコン整えるとかそういうような話です。

〇懸念されるリスクの具体例と対応

懸念されるリスクの具体例と対応については、次の６点があります：１）機密情報の漏洩や個人情報の不適正な利用のリスク 、２）犯罪の巧妙化・容易化につながるリスク 、３）偽情報などが社会を不安定化・混乱させるリスク、４）サイバー攻撃が巧妙化するリスク 、５）教育現場における生成ＡＩの扱い 、６）著作権侵害のリスク、７）ＡＩによって失業者が増えています。

澤井は、日本の場合むしろチャンスにすべきで、ＡＩによって生産力を上げ、逆に職業が増えていくという風にすべきだと思っています。

〇ＡＩの利用

ＡＩの利用については、次の４点が考えられます：

１）生成ＡＩは、デジタル化を加速させ、我が国全体の生産性向上のみならず、様々な社会課題解決に資する可能性があります。２）ＡＩ利用を加速するため、医療や介護・行政・教育・金融・製造等のデータ連携基盤の構築・ＤＦＦＴ（Data Free Flow with Trust）構想の具体化・人材育成・スタートアップの事業環境整備を進めるべきです。３）政府機関が一体となって、機密情報漏洩のリスクなどに配慮しつつ、率先して生成ＡＩの利用可能性を追求することが重要です。４）幅広い世代が生成ＡＩの恩恵を享受できるよう、スキル・リテラシーを身に付けることが大切です。
ＡＩの利用は、デジタル化を加速し、我国全体の生産性向上のみならず様々な社会課題の解決する可能性があるので、教育に関しては幅広い世代が生成ＡＩの恩恵を享受できるスキル・リテラシーを身につけることは大切ではないかということです。

・生成ＡＩを活用した授業の実践例

【事例１】千葉県印西市立原山小学校（松本博幸校長）の生成ＡＩを活用した授業の実践例です。授業では「文章生成ＡＩの仕組みを知り、これからどのようにかかわっていけばよいか考える」ことを狙いとし学習活動に取り組んでいます。保護者の理解及び協力、そして専門的な知見を有する外部人材の助言によって、全国的に見ても先進的な生成ＡＩについての授業が実践されていました。
この授業実践例では、専門家の「ＮＰＯ法人　みんなのコード」が「対話ＡＩは、ネット上にある大量のデータを使って、それをまずＡＩに学習させて、人間の質問や指示に答えるようにするのだよ。悪いことに使わなければうまく使うことができるよ。ＡＩの言っていることは１００％正しいということはないということを理解して使うのだよ。」とわかりやすく教えています。

【事例２】同志社中学校・高等学校 の反田 任先生のＡＩ発音チェックアプリ”ELSA Speak”を活用した英語授業デザインの事例です。 目的は ELSA Speak を活用し「話すこと（発話）」の力を向上させることです。ELSA Speak では英語の音素単位で発音判定をするため、単語や英文の正確な発音が身に付きます。

・生成ＡＩが脅威的な進化

生成ＡＩが脅威的な進化をしています。最近「ＡＩアライアンス」が出てきました。ＡＩアライアンスとは、Microsoft の「being　ＡＩ」、ChatGPT、GPT4 や GPT4 ターボに対抗する国際的な連合グループです。

・雇用の未来

オズボーン氏の論文「雇用の未来」の中では、 「２０３０年には、現在ある職業の４７%がなくなる」と言われています。そのため、ＡＩ時代を生き抜く人材を育成する教育が重要になる。消える職業、なくなる仕事で筆頭に上がってるいのは銀行の融資担当者とかスポーツの審判、不動産のブローカー、レストランの案内係や、保険の審査担当者というようなことが上がっています。

３ 自動運転

自動車産業領域では、ディープラーニングが強みを持つ「認識」領域を中心に、数年前から様々な実証実験が行われています。安全性を最重視しつつ、将来の実用化に向けた取り組みが進められています。

・自動運転

カメラ・センサーなどから「信号・道路標識・障害物」等の外部情報を把握し、アクセスやブレーキ、ステアリング等の各種操作につなげていく自動運転の工程内で、ディープラーニング技術が活用されているわけです。
実際の動きを見てみますと、車の上にセンサーが付いていて、他にもセンサーついていると思うのですけど道路の障害物とかを認識して動いています。
内閣官房ＩＴ総合戦略室等では、高遠道路でのトラックの隊列走行を早ければ２０２２年に商業化することを目指しています。高度自動運転の市場化・サービス化にあたり、様々な走行環境における実証実験の実施」が不可欠となって来ました。
成果としては、まず自家用車の分野では２０２１年３月にホンダより、自動運転レベル３に適合する技術であるトラフィックジャムパイロット（渋滞運転機能）を搭載した車両が市場化されました。
ホンダの自動運転レベル３は、高速道路渋滞時など一定の条件化で、システムがドライバーに代わって運転操作を行うことが可能になりました。

このことにより、２０２０年の実現を目指していた高速道路での自動運転レベル３が実現されることになりました。また「運転支援システムの高度化」について、高精度３次元地図を利用した速道路でのハンズオフの運転支援が日産などで実現されました。

自動運転の国内外の動向を見てみましょう。
アメリカの場合、民間主導でレベル４を実用化しています。具体的には、ＩＴ企業が自動車を時期デジタル領域と見据えて、無人モビリティサービスの事業開発と促進しています。ＧＯＯＧＬＥなどのＩＴ企業を中心に、Ｗａｙｍｏタクシーや Ｇａｋｉｃ無人配送トラックといったレベル４サービスカーが一部公道で実用化しています。
欧州では、欧州ＯＥＭのディーゼル不正やＨＥＶ技術力不足を補う競争力獲得のため、自動化・電動化を政府指導で推進しています。官民ともに、レベル２のとかレベル３が中心で、官主導でレベル４の試みが行われています。レベル３までは車の運転を基本的に人間が主導で行っています。レベル４以降では基本的に人間はアシストしません。レベル５に至っては人間のアシストが全くありません。今見る限りレベル５は、未だ出てきていません。
中国の場合、２０２５年までに製造強国の仲間入りを、政府目標としており自動化技術開発を政府指導で推進しています。
新興国では、民間主導でレベル４を目指しており、公共交通の未整備地やスマートシティ内の移動手段確保を目的に、民間を中心に自動運転を実証しています。

日本における自動運転の導入地域とＡＤ／ＡＤＡＳ市場の拡大です。ＡＤは「自動運転システム（Autonomous driving system」で、ＡＤＡＳは「先進運転支援システム（Advanced Driver-Assistance Systems）」です。
自家用車については２０２２年にレベル１からレベル２までに普及しています。２０２５年には運転支援の進化により、レベル２からレベル３になりＡＤ／ＡＤＡＳ導入が促進されます。そして高度な運転支援車とサービスカーの基盤を共用し、ＡＤ／ＡＤＡＳ　ＥＣＵ、センサーやアルゴリズムなどの量産効果によりコスト低減されたシステムの活用を通して、商用車レベル４サービスカーの本格普及に寄与します。
商用車については２０２２年にレベル４のサービスカーを導入し、自家用車との協調領域における自動運転開発を行いながら、２０２５年にはレベル４サービスカーの導入を４０箇所に増やし促進して行きます。２０３０年以降に自家用車の技術を活用しながら、レベル４サービスカーの本格普及を目標にしています。

自家用車のレベル４自動運転サービスの実現は、限定空間の遠隔監視のみで自動運転サービスの実現に向けた実証事業の推進という形で行われています。事業性向上に向けて、４台までの車両を１人が同時監視します。２０２３年5月２１日より、福井県永平寺町において、レベル４での自動運転移動サービスが開始されました。西村経済産業大臣（当時）が乗って現地で記念式典に臨んでいます。
町民の方々による乗車や、遠隔監視室でレベル４運行を監視して様子が見れます。遠隔監視室では、１人の人が３台の自動運転を監視し、操作しています。

・タクシー需要予測

タクシーの需要予測にもＡＩが使われるようになりました。タクシードライバーが抱える一般的な課題に、「需要がある場所や時間帯を予測するのが困難」というものがあります。これから紹介するケースでは、タクシー需要を予測するために各種データ(リアルタイム人口統計データ、気象データ、過去のタクシー運行データ等)を活用して、どこに車を配車したら１番効率がいいか正解を予測する本事例では、「抽象的で複雑な特徴を獲得する」ことを狙い、オートエンコーダ系技術を採用しています。オートエンコーダとは、３層ニューラルネットにおいて、入力層と出力層に同じデータを用いて教師あり学習させたものです。
【事例】株式会社ＮＴＴドコモの「ＡＩタクシー」は現在から３０分後までの未来のタクシー乗車需要の予測結果などのデータをオンラインで配信するサービスです。需要を予測するために、発展させたオートエンコーダ(Stacked denoising Autoencoder 、ＳｄＡ)を使用しています。タクシー需要を９２.９%の精度で予測しました。実証実験を経て、２０１８年2月よりサービス提供開始し２０２２年6月に終了しました。

・未来の自動運転

２０２３年8月１３日、米国ミシガン州政府は、デトロイトからアナーバーの間に全長４０マイル（約６４ｋｍ）に及ぶ米国初の自動走行レーンを建設すると発表しました。レベル5の自動運転は難しすぎるというので、アメリカでは道路になんか仕掛けをしようというようになってきているようです。デトロイトというとアメリカの自動車の発祥の地なので、デトロイトで行うことが将来の自動運転に相当のインパクトを与えるわけです。
これまでグーグルは、インフラに依存しない「自立型」の自動走行の実現を目指してきたことで有名です。

２０００万マイル（約３２００万ｋｍ）以上の公道走行を続けながら開発してきた技術です。２０１６年に設立された自動運転専門会社「Ｗａｙｍｏ」が、アリゾナ州で提供している有料の自動走行ライドシェアサービスＷａｙｍｏ Ｏｎｅのいずれも自立型の自動走行を前提としてきました。
グーグルは、関連会社Ｃａｖｎｕｅがインフラ協調の必要性を主張したことにより自動運転の考えが変化しつつあるようです。
オレンジに引いた線のところが、自動運転車レベル５が走るイメージです。混雑している道路の中でオレンジのところだけは自動運転で走っています。

４　国際的なＡＩのルール作り

上図は、国際的なＡＩのルール作りです。政府は「ＡＩ戦略２０２２」や「人間中心のＡＩ社会原則」などをこれまで定めてきました。

論点整理は、従来の基本戦略・理念は維持しつつ、急激な技術変化やＧ７広島サミットで合意されたビジョンと目標（「我々が共有する民主的課に沿った、信頼できるＡＩ」）等を踏まえ、ＡＩ戦略会議の構成員が有識者として、生成ＡＩを中心に課題と方向性などを整理しています。

・広島プロセス

広島プロセスには３つの動きがあります。
１）２０２３年５月Ｇ７広島サミットにおいて岸田総理が提唱したことです。
２）広島ＡＩプロセスのＧ７デジタル技術・閣僚声明です。広島ＡＩプロセス包括的政策枠組みです。①生成ＡＩに関するＧ７の理解に向けたＯＥＣＤレポート、②全てのＡＩ関係者に向け及び高度なＡＩシステムを開発する組織向けの広島プロセス国際指針、③高度なＡＩシステムを開発する組織向けの広島プロセス国際行動規範、④プロジェクトベースの協力の４つあります。広島ＡＩプロセスを更に前進させるための作業計画の策定です。
３）２０２３年１２月６日Ｇ７首脳による最終合意で、ＡＩアラスシンポジウム２０２３で公表された指針です。①全てのＡＩ関係者に向けの広島プロセス国際指針として、導入前及び市場投入前を含め適切な措置を講じます。②導入後の話です。③十分な透明性の確保を支援します。④関連組織間での責任ある情報共有します。⑤ＡＩガバナンス及びリスクの管理方針を策定し、実施し、開示します。⑥強固なセキュリティ管理に投資し実施します。

⑦ユーザーがＡＩを生成したコンテンツを識別できるようにします。⑧社会的、安全、セキュリティ上のリスクを軽減するための研究を優先します。⑨世界の最大の課題、特に気候危機、世界保護、教育等に対するために、高度なＡＩシステムの開発を優先します。⑩国際的な技術企画の開発を推進することです。⑪適切なデータインプット対策を実現し、それによって個人データ及び知的財産を保護します。⑫高度なＡＩシステムの信頼でき責任ある利用を促進し、デジタルリテラシー、訓練及び認識を向上させる機会を求めるべきであるということです。

・リスクの対応として

中央省庁が生成ＡＩを利用する場合の手続きで、生成ＡＩの業務利用に関します。２０２３年９月オープンＡＩに対する注意換気がありました。
ＡＩと知的財産権とＡＩと著作権問題に関して、研究会・審議会における議論が行われています。教育とも絡む面では学習用データの整備で、政府等保有データの提供促進があります。加えて、新たなモデルの開発、基礎的・先進的な研究、人材育成があります。注目したい点は次の３つあります。

１）ルールメイキング＋エンフォースメント、２）大規模汎用モデル＋多様なモデル、ファインチューニング、３）一部の地域での利用＋世界各国での本格的な利用というものがあります。今後、安全性・信頼性、オープンクローズ、多様性/国際整合性の議論がますます重要になります。

５ ＡＩ倫理

現在の第４次ＡＩブームでは、自動運転や画像診断など私たちの暮らしにＡＩ技術が急速に入り込んで来ています。２１世紀の基幹テクノロジーとされるＡＩとどう付き合い、その活用をどこまで許容していくのか？ＥＵではＡＩ倫理に基づく輸入規制を計画しており、日本のＡＩ倫理が問われています。

・ＡＩ倫理の背景

２０２３年１２月１７日ＯｐｅｎＡＩの「超知性」誕生に備える研究チームがＧＰＴ－２（弱いＡＩ）モデルでＧＰＴ－４のように強力なＡＩ（強いＡＩ）を制御する方法を説明しました。ＯｐｅｎＡＩは、人間よりもはるかに賢いＡＩである「超知性」が２０３３年までの１０年間で開発されると推測しており、「スーパーアライメントチーム」を立ち上げ、超知性を制御するための研究が行われています。

ＡＩの賢さを下回る人間ではＡＩの監視が困難になります。ＯｐｅｎＡＩのスーパーアライメントチームは、人間が超知性を適切に監視できるかを見る代わりに、大規模言語モデルのＧＰＴ－２がより強力なＧＰＴ－４を監督できるかテストを行いました。

注）ＧＰＴ－２のパラメータは１５億程度であるのに対し、ＧＰＴ－４のパラメータは約１７６０億に上るといいます（ＩＥＥＥ　Ｓｐｅｃｔｒｕｍ）。

・ＡＩ倫理の定義

倫理とは、Ｗｅｂｓｔｅｒによれば「ａ　ｓｙｓｔｅｍ　ｏｆ　ｍｏｒａｌ　ｐｒｉｎｃｉｐｌｅ」となっており、ＡＩ倫理は「ａ　ｓｙｓｔｅｍ　ｏｆ　ｍｏｒａｌ　ｐｒｉｎｃｉｐｌｅ　ｆｏｒ　ｕｓｉｎｇ　ＡＩ」と定義できます。日本の文部科学省が推進する全国の児童・生徒１人に１台のコンピュータと高速ネットワークを整備する「ＧＩＧＡスクール構想」に必要な「ＩｏＥＥＥ」（倫理のインターネット、 教育のインターネット、生きる力のインターネット)ＡＩ倫理チャットボット機能を試作・検証することにあります。

・１人１台端末に必要なＡＩ倫理チャットボット

ＡＩ戦略の教育改革「１人１台端末」、ＧＩＧＡスクール構想の最大の課題は「チャットによるいじめ問題」と言われています。ＡＩ倫理を探究するうちに「チャットによるいじめ問題」にＡＩ倫理チャットボットが使えるのではないか？と考えらています。

近年、人間と会話をすることができる対話システムへの注目が集まっています。例えば、音楽の再生やメールの確認などを行う Ｇｏｏｇｌｅ　Ａｓｓｉｓｔａｎｔ や Ｓｉｒｉ、また顧客からの問い合わせ対応を代替するチャットボット（チャットのロボット）といった、何らかのタスク達成を目的としたタスク指向型対話システムが広く浸透してきています。今日、自動運転や画像診断など私たちの暮らしにＡＩ技術が急速に入り込んで来ています。２１世紀の基幹テクノロジーとされるＡＩとどう付き合い、その活用をどこまで許容していくのか？「ＡＩ倫理」とでも呼ぶべき社会規範をきちんと議論しなくてはならないと言われています。

ウイーン大学の哲学者クーケルバーク氏は著書「ＡＩ倫理」で、ＡＩを使うための「運転免許証」がない。と警鐘を鳴らしています。

ノーベル文学賞受賞者のカズオイシグロ氏は「クララとお日さま」と言う最新の小説の中で、人工知能を搭載したロボット「クララ」を登場させています。クララは、観察と学習への意欲と理解力を持つに至り、人間社会で生きていく力「生きる力」(Ｅｎｅｒｇｙ　ｏｆ　Ｌｉｆｅ)を得るようになります。

・教師あり学習

機械に学習させる「機械学習」には「教師あり学習」、「教師なし学習」、と「強化学習」の三つの学習の枠組みがある。図５は人間の脳のニューロンが層状に接続した構造を模擬した機械学習の三つの枠組みです。
教師あり学習とは主に人間の小脳が担う学習機能で、代表的な統計手法は回帰と分類です。学習者に対し、教師が明示的に正解を教えたり、学習者の誤りを指摘したりすることで、学習者が正しい解を得ることを助けます。正しい入出力の組合せを与えて学習することで、新規の入力に対し適切に出力します。
「回帰」の代表的手法は誤差逆伝播法(Back Propagation)です。「分類」の手法として、正解、若しくは誤りを入力として、未経験入力に対する意志を決定する決定木(Decision Tree)や決定表(Decision Table)の作成などがあります。本研究では EXCEL 上の決定表で「倫理表」を試作しました。

・「ＡＩ倫理」処理システムの試作

「教師あり学習」ＡＩを使い、社会規範・倫理と、設計者の故意ではないＡＩの誤認識（機能不全、誤作動や機能低下を含む）を検証し、適切な処理を行う「ＩｏＥＥＥ」（Internet of Ethics, Internet of Education, Internet of Energy of Life)ＡＩ倫理チャットボット機能の試作を行いました。
本研究では、「教師あり学習」を使い、教育禁止用語や放送禁止用語等のような社会規範・倫理とＡＩの誤認識が処理・説明できるシステム作りを目指しました。入力はＡＩ音声入力でもキーボード入力でもできます。デープラーニングによるＡＩ音声入力は iPhone で行い、リモートマウスで接続したパソコン上でＡＩ倫理処理を行いました。「ＡＩ音声入力では何故誤認識したか？」は言葉では説明できない、つまり暗黙知です。
社会規範・倫理とＡＩの誤認識の検出・修正（言換え）処理は EXCEL の VBA プログラムで瞬時に処理され、修正した音声入力文と修正理由を説明した説明文はそれぞれ EXCEL ファイルに保存される。学習データは、社会規範・倫理例、ＡＩの誤認識と学習済みの TensorFlow.js モデル・デーモン（システム）等で、インターネットとブロックチェーンで参照します。

例えば入力文「Slave is a bad word」（奴隷は良くない言葉です）を入力すると、正しい表現に言い換え、その理由を説明する社会規範・倫理例の放送禁止用語は教育禁止用語としてウェブ検索する具体的にはアイヌ系からロンパリに始まりブスとかチビといった誹謗中傷の類からジョンやアメ公といった人種差別用語まで教育上使わない方が良いとかんがえられる用語は網羅されています。

・「教師あり学習」モデルを使った検証

音声入力文に、①アイデンティティベースの憎悪、②侮辱、③わいせつ、④重度の毒性、⑤性的に露骨、⑥脅威、⑦毒性などの有毒なコンテンツが含まれているかどうかを、約 200 万件を事前に「教師あり学習」した学習済みの TensorFlow.js モデル・デーモンを使い検出しグラフ化し、「ＩｏＥＥＥ」ＡＩ倫理チャットボット機能の検証を行いました。例えばＧＩＧＡ端末の入力文「馬鹿!消えてしまえ！」を、学習済み TensorFlow.js モデル・デーモンに入力し分類すると、②侮辱）かつ⑦毒性が「ＴＲＵＥ（きわめて有害）」、及び①アイデンティティ攻撃、③卑猥、④重度の毒性、⑤性的な露骨及び、⑥威嚇は「ＦＡＬＳＥ（無害）」と分類します。
ＧＩＧＡ端末でトラブルを引き起こす入力文例 37 件中 26 件(10%)がＴＲＵＥ(きわめて有害)、53 件(20%)がＮＵＬＬ(要注意)、残りはＦＡＬＳＥ(無害)と検出できました。「ＩｏＥＥＥ」ＡＩ倫理チャットボット機能は、30%を超える抽出率で、倫理テーブルで確認できなかった未定義の誹謗中傷を検出できました。本チャットボット機能は 非常に効果的でした。

結果、ＧＩＧＡ端末やケータイに人工知能を搭載した人間に親切な「ＩｏＥＥＥ」ＡＩ倫理チャットボット機能を搭載し、１）社会規範・倫理とＡＩの誤認識の修正処理を行い、その後２）「教師あり学習」モデルを使った検証を行うと、抜けが少ない有効なＡＩ倫理処理ができると分かりました。

課　題

自動運転車が引き起した事故は、誰が責任を負うべきか？について考察し，あなたの考えを 800 字で説明しなさい。