

文化勲章受章記念特別講演 長尾真（京都大学名誉教授・元総長、日本学士院会員、
国際高等研究所学術参与・前所長）

「AI は哲学の最前線」

みなさんこんにちは、長尾でございます。昨日今日と大変刺激的なお話を皆さんからお聞きしまして、大変感銘を受けましたし、わたくしは 1959 年に学部を卒業しましてから、文字読み取りの研究これは郵便番号読み取りに利用されて。その後画像処理の研究を一生懸命やりまして。これは人間の顔画像の認識の研究で、世界で初めて人の顔のアイデンティフィケーションというのをやったんですけれども、先ほどおはなしがありましたけれども、人の顔を分析的に研究するというのは非常に面白いんですけれども、これは人間の心を見通すような研究に繋がっていく可能性がある。そうするとこれは人間の尊厳というものに非常に関わるという気がいたしまして。そこで私は研究をやる面白いんだけどやらないという自制をして、やるべき研究、やるべきでない研究というのを意識しまして顔画像の認識はその後研究しなかったんですけれども、一般的に色々な人たちが研究をやりまして現在の街角にある監視カメラなんかはもうものすごく発達して泥棒とか色んな人が行き来するのを全部個人単位で認識できるようなところまでやってきた。これがいいのか悪いのか。社会の安定につながる可能性はあるけれども、個人のプライバシーとかそういう問題との微妙な関係があるというような問題も喚起してきているというそういう微妙な時代になってきたことがございます。で顔画像の認識を辞めまして、そのあと言語情報処理の研究をずいぶんやりまして。そして日本語と英語の間の機械翻訳の研究をやりまして 50 年くらいやって今日ディープラーニングの技術を取り込むことによってほぼ実用の性の高い機械翻訳システムができるようになってきたわけでございます。これは大変けっこうなことで今 30 か国以上の言語がかなりのレベルで翻訳できるということになってきました。あと 10 年 15 年やれば装置も安くなるし、いろんな世界中の人たちが機械翻訳システムを使って相互交流ができるあるいは異民族間の相互理解が促進されるとそういうところになっていくんじゃないかと。それでできれば世界平和に貢献ができるといいんじゃないかとそういう風に思ったりしているところでございます。

でそういう関わりをもって人工知能というものを昔から考えてきたんですけれども、結局人工知能というのは知能というものを人工的に実現するというまさにその言葉通りでございます。人間頭脳の活動をいかにうまくシミュレーションして人間が考えているのと同じようなことを機械にやらせることができるかという、そういう研究であるということになるわけでございますので、人間の頭の中でどういうことが行われているか、

特にものの考え方、人間のもってる思想というものがどういうものであるか、それが機械の上で本当にどこまで実現できるかということに関心をもっているわけでございます。そういうことを考えますと、古代ギリシャの哲学から今日までヨーロッパ思想というものがどういう風に発展してきて、どこまで来ているか、そういうものが人工知能にどこまで役に立つか。あるいは逆に人工知能で今やっている研究というものが哲学の世界にどういう風に取り込んでいってもらえるか。そういうことを色々論じてみたいと思っているわけでございます。ですから今日はそういうことを中心にお話して、そしてもう一つは人工知能が社会にもたらす問題点、そういうことについて最後に少しお話しをしたいと思います。

ということで哲学というものはこれまで何を論じてきたかというところとやっぱり宇宙の成り立ちあるいは宇宙とはいったいどんなものであるかという宇宙の構造、そういうものに古代ギリシャの人たちは大変関心をもった。それからもう一つは人間とは何か。ソクラテスを始めプラトンやアリストテレスは学問を築いたわけでございますけれども、人間とは何かというのがそのスタートポイントであった。で、これが流れ流れて人工知能の世界にまで来ると考えていいのではないかと思います。それからもう一つは、存在とは何か。人間存在は当然ですけども、人間だけでなく世界というものの存在それがいったいどういうものであるかということに関する根源的な質問、それにその理解をどうするかということを考えてきたわけでございます。で、簡単に言いますと哲学は頭脳の根源的な働きに向かって進んできた、根源的なものというのとは何かと言うと、魂みたいなものなんですけども、最初は脳皮質の働きを主に研究したわけですね、つまり論理の世界。これはプラトン、アリストテレスからずーっと下がってきてカント、ヘーゲルくらいのところまでは脳皮質つまり人間の持つてくる論理の世界そういうものを徹底的に調べ議論するところをやってきたわけでございますが、それがヘーゲルで一段落をしまして、哲学がある程度完成したと思ってたら、キルケゴールが、いやそんなものでは人間というのはわからないんだと。人間というのは心があって心が苦しんでる。心の存在というものを考えなければ人間の存在というものは意味がないんだという所から始まって、人間のもってる心つまり感性とか感情とかそういう心の世界を徹底的に研究するということが始まったわけでもメルロポンティからもうちょっと後までぐらいがそういうことに関する哲学世界における議論のさかんな時代でございましたけれども、一方ではフロイトが無意識の世界というのを発見して、人間というのは論理の世界あるいは感情の世界とかそういうものだけでなく、気づいていない脳の奥底には何らかのものがあると。それが夢に出たり色々する。つまり人間の頭脳のもってる記憶の世界というのがどういう構造をしているかということについて疑問を投げかけた。これが非常に面白いという人工知能にもものすごく関係をするという風に私は理解しております。ピアジェなんかも集団の心の動き、あるいは集団の持つてくる無意識そういうものまで考えた

わけでございますけれども、そういう活動をふまえて、ウィトゲンシュタインなんかが代表的ですけれども、言語哲学、つまり哲学が色々論じられてきたけれどもそれは論理の世界だけではなくて結局それは人間として言葉でそれを説明して言葉でコミュニケーションをして相互理解する。そういうことから言葉というものは一体哲学的に考えてどういうものかということを考えなければ哲学の世界を正確に把握することはできないというところから言語哲学がさかんになったわけでございます。でそういうものがアメリカのほうへ流れて行って、あるいはナチスの圧迫とかいろいろありまして、ウイーン学派の人達もアメリカへ行ってク、ワインとかいろいろな人達が結局分析哲学というところに行ったわけでございます。この分析哲学ってのはなかなか把握することが難しいと思いますが、簡単に言いますと哲学議論というものをできるだけ明晰にやりたいと、曖昧性をなるべく少なくしながら哲学内容、思想の内容というものを伝えたい。ということから、論理学その他を使いまして、できるだけ明確に議論をするという立場で哲学的諸問題を論じるというところに行った。これがアメリカの人たちがやってきた分析哲学でございます。一方ではフロイトの無意識の世界から発展してきました心理学であるとか、あるいはそれが自然科学的な技術、tool を用いることによって認知科学の世界というものが開けてきたと言ってもいいんじゃないかと思えます。でそこまで来てじゃあその次はどうなるかということそれが人工知能の世界に結びついてきてるんじゃないかというのが、私のこじつけみたいなことかもしれませんが、考え方です。つまり人間の頭脳の活動というものを具体的に知るためには、コンピューターの上で頭脳の働きをできるだけ正確にシミュレーションするということが人工知能の一つの大きなミッションであるということかと思ってます。それからその次じゃあそれが上手くいったらどうなるかと言うと、それは昨日入来先生がお話になった非常に魅力的な講演にありますように、脳機能というものを神経科学的に解析する、説明する、そして神経科学的な脳の働きがシンボルの世界、抽象世界とどういう風に結びついていくか、そういうことをしっかり押さえていくというのが人工知能のつぎの学問的發展の世界ではないかというのが私の考え方でございます。で、そういうことになぜなってきたかということ、全ての学問には手段というもの、つまり使う tool というものが必要であるということがいえるわけですが、古代ギリシャの哲学は対話というものを一つの tool として議論したわけでございます。でそのあとはやはり書物というものを通じて思考を深めるということが行われる。つまり書物というのが思想の表現の一つの tool になったというわけですね。それからその次は実験あるいは観測というものを通じて人の心の内容を知るという、フロイトとかそういった人たちの手段、これは無意識の世界は哲学の世界ではないという意見も多いでしょうけども、哲学から発展した世界と見ることもできるわけですし。実験とか観測というものを通じてできるだけ客観的にとらえていこうとこういう風になって。でもう一つの分析哲学とか言語哲学みたいなものは論理学などを用いてできるだけ

客観的に議論をしたいということが行われてきたわけでありますが、その後に出てきたのは今さっき言いましたようにコンピューターというものをツールとして議論をする、それが人工知能であって哲学の議論の最前線にあるのが人工知能だという風な大げさな議論を今日はさせていただくことにしたわけでございます。

そこで繰り返しになりますけども、人工知能っていうのは人間頭脳の働きについての実証的な哲学だという風に見てもいいんじゃないかということでございまして。現在の AI は大脳皮質のレベルでの人間頭脳の活動を実現しようとしている。心の問題っていうのはもっともっと先で、現在の AI は大脳皮質の働きを模擬するレベルにとどまっている。だけどそれはどういう内容かという色々なセンサーによる入力を認識する。人間には五感というのがあって目とか鼻とか口とか耳とかあるいは触覚とかいろいろある。そういうところから入ってくる外界の情報をどういう風に認識するかという問題、それから推論をする、それから知識を使う、それから学習をする、とこういうのが基本的に最重要の機能だと思いますけれども、その中でやっぱり一番これからの人工知能の研究にとって大事なのは知識をどういう風にコンピューターの中で蓄積するか、そしてそれが誰でもがうまく使えるように知識をオーガナイズする。そういうことについては全くバラバラになってなっちゃってしまってる。データなんかは皆さん方の議論の中にありましたけども、フォーマットが違くとユニバーサルになかなか使えないということがありますね。

それに対して推論の機能というのがあるわけですが、これをどうするかというのが大問題ですね。人間の頭のなかで脳神経細胞が色々働いてるけど、その働きが推論に対応している。その推論、コンピューターでやると推論の段階がいくらかでも深くやっていける。それに対して、人間の頭の中での神経細胞がやってる推論は、たぶん 5・6 段階ぐらいやるとそれ以上もうとてもできないようなことになってるわけですね。推論というの、ある一つの方向にむけて推論するだけじゃなくて、コンピューターによる囲碁や将棋がなんで強いかというと、あらゆる方向に推論をして、その中から一番有力なものをチョイスする。そういうことができるようになって 5 段階 10 段階もっとも深く推論することができるわけですね。コンピューターにやらせると。ところが人間なんかの場合は相当の名人でもある方向だけだと深く推論して 50 段階ぐらいはやれる人がいるらしいですけども、あらゆる可能性についてそこまで深く推論することが人間の頭脳ではできない。そういうことがありますね。それから推論していくときに一番大事なのはやはりどういう知識をたどって推論していくかということがあるわけなんで。知識というものを徹底的に集めてそれを使いやすい形にシステムティックに整理をするかという問題があるわけで、それについては、人工知能はまだまできていない。人間の頭の中で知識がどういう構造で存在しているかというのもほとんどわかっていない。こういうことになるわけですが、応用的に考えるとコンピューターにおける知識という

のはある意味でかなり使えるものが出てきているわけですね。囲碁将棋なんかはその例ですけども、その他に裁判なんかの判例とかそういうものでも知識が相当たまってきている。そういうことになっていまして、コンピューターのパワーでもってかなりのことができる、という風になってきたわけでありまして。学習機能をもつことによって人工知能は飛躍的に進歩しているわけでありまして、そのことを考えると、大脳皮質の働きのレベルにおいては人間は人工知能には勝てない。今は勝ててるかもしれないけども、あと 10 年もすれば人工知能に人間の頭脳は大脳皮質の働きのレベルにおいては勝てないというのがわたくしの見解であるというか、そういう感じをしております。古代ギリシャから今日まで延々と思想が流れてきたわけでございますけれども、結局人工知能を徹底的にやれば人間はたぶん勝てないんじゃないという風に考えているんですね。残念ですけど。

ただ、考えようによってはですね、たとえば 360 度あらゆる方向に推論をしていく。人間もそりゃあやればできないことはないと思いますけども、その推論した結果いろんな結果が出てくる、その中で何をチョイスしてそれを最終決定にするか。囲碁、囲碁や将棋でいえばどの手を打つかということを決めるためには推論した結果がものすごくたくさんある中でどれか一つをチョイスする。それがどういう風にして行われているかということを見るとそれは人間の場合は個人個人のもってる価値判断に拠ってるわけですね。価値判断というのはどういうものであるか。これはなかなか難しいんですけども、やはり決定というのは価値判断を基づいて可能性の中から一つをチョイスすることになるわけですね。そうすると、人間の場合は価値判断というのはどういう風にして育成されてくるかという問題が心理学上非常に面白い問題であります。いずれにせよ人間は価値判断力をもってる。それが個人個人のパーソナリティを形成しているということになるわけですけども。じゃあコンピューターは価値判断力をもってないか。今の段階ではプログラマーが価値判断力を人工知能のシステムに与えてるわけですけども。人工知能が発展していくと、自分自身で価値判断力を形成していくことができるのかできないのかという問題が出てきます。で私はそれは自動的にできるという風に思っております。プログラムをうまく作ることによって AI システム自身が、自分の価値判断基準をつくる。たとえば、なぜそんなことを私が言うかといいますと、人工知能があるチョイスをして仕事をしたというときに、それを見ている人間が「うまいうまい」というか「そんなのダメだよ」というかというのを、人工知能自身が見て判断する。そういうことを何千回何万回という風に繰り返していくと、人工知能がやるチョイスのシステムというのが自ずからつくられて行く。人間が「いいよ。このやり方はいいよ」と言ってる方向に人工知能の価値判断力のシステムが形成されていくと。そういうことが十分あり得るんです。このように価値判断のシステムも人工知能が自分で作り出すということが出来るんじゃないかと思っております。そういう意味でも、人間は人工知能

に勝てないと言えるのではないかというわけでございます。色々と反対があると思いませんけども。

でじゃあ人間は一体人工知能にどういう面で勝てるんかというか、勝ち負けの話ばかりしてるとしょうがないんですけども、人間の存在価値というのはどこにあるんだということを考えますと、それはこの論理の世界だけでなく感情などの心の世界の解明。心の世界というのを人間はもってる。それを人工知能は今の段階では少なくとももってない。もっていませんね。それをもてるかどうか。もてるとしたらどこまで、どの程度までもてるかということが問題になる、ということになります。そこで AI は人の心をどこまで理解できるかという課題が出てくるわけでありまして。これが大きな問題です。で現在は感性であるとか感情というものをある程度持つことができる、機械的にですね。そういう研究がずいぶん進んできておりますので、感性、感情、人間の心（人間の心というのは何かという定義を与えないで話をしてるのはあまりよくないんですけども、それをやりだすと時間がありませんから、やめときまして）、そういうものについて、ある程度はやれる。だけど、まだまだ完全じゃない。そういう中で感性・感情とかいうのがどうして把握できるかというときに、言語というものを一番よく考えなければいけないんじゃないかということでありまして。人間の表情というのはもっともっと研究すれば、いろんなことがわかってきます。それとおんなじように、言語表現、人の発話というところに含まれている心の表れというものをどこまで把握できるのかということが問題になるのですね。で現在日本だけじゃなくて、どこでもやっているロボット研究というのは残念ながらですね、メカニカルなところはかなりよくできるようになってきたんですけども、ロボットの頭脳の中は殆ど空っぽですね。それが一番の問題でしてロボットの頭に言語を理解する能力をもたせる、あるいは人間の表情をちゃんと理解することができるようにする、そういうプログラム、といったものを入れないと意味がない。だからこれからは人の心の表れ、それがビジュアルに現れる場合と音声的に現れる場合と言語という世界で現れる現れ方と色々あるわけですけども、そういうことについて徹底的に研究するということが一番大事になる。AI の近い将来の課題というのはそういうところにあるのではないかと思っています。たとえば今一番問題になってますのは介護ロボットをつくらないといけないというのは世界中だれでもわかってきておりますし、日本は特に高齢化社会ですから介護ロボットをつくらないといけないんですけども、その介護ロボットは介護される人の心をちゃんと把握できるように、そういうものでないと意味がないということになりますので、何が何でも人の心をちゃんと把握できるような介護ロボットをつくっていくことが介護の世界だけでなく、他のあらゆる世界において必要であるということになります。

そういうことをロボットの研究開発については私が考えておりますが、もう時間がなくなりましたので、それでは人工知能と社会の問題について簡単にお話したいと思いま

す。人工知能が発展発達すると人の仕事が奪われると、つまり失業がふえるということが言われてるわけですね。たしかにそれは事実そうなると思うのであります。大脳皮質、人間の大脳皮質の働きだけで行われているような仕事は殆どがロボットに変わる。で人間の心が重要な働きをするような世界に関してはなかなかロボットは簡単に入るところまではいかないので、これは人間の仕事になる。そして創造的なクリエイティブな仕事は人間がやるべきだということになるかと思ひます。いずれにせよ失業がふえるから、ベーシックインカムという考え方で色んな事を考えないといけないというのが、今騒がれているわけですね。で私が言いたいのは、ロボットを導入することによって、人間よりパフォーマンスがよくなりますから、それによって得られる利益の一部を税金でとって、それをベーシックインカムとして失業する人にわけるとというのが一つの方法だと思ひます。しかし、それよりももっとやるべきなのはワークシェアリングをやる。つまり一つの仕事をやるのにひとりの人が8時間働いたらできる仕事だけでも、失業者がいるということであればそれはよくないので4時間4時間にして失業してる人呼び込んで仕事を一緒にやるというようなことをやることによって、失業をなくしていく。つまり、人間というのはやっぱり仕事をするというか働くというか、体を動かす、そういうことをやらないと健全な人間ではなくなっていく可能性があるわけですから。ベーシックインカムをもらって生きていくというのはそういうのは人間としてはまことに情けない状況なので、ワークシェアリングをやって、健全な人でも半日働く。あとの半日は自分の好きなことをやる。収入は半分に減るかもしれない。しかしそれはベーシックインカムのためにとった税金の一部を与えるというような形なるべく失業者を減らす。そういうことはやればできることであるという風に思ひます。たとえば、京都のいろいろな仕事をしている人たち、たとえば帯を作るとかですね、いろいろな仕事をしている人たちが一人ででもできるんだけど、それを3人も4人もに分担して分けて少しずつやるっていうことによって、多くの人がある仕事に携わって失業者を出さずに、そして適当な、金はもうからないかもしれないけども、生きていくだけの収入は得られるというような、そういうことはずっと昔からやられてきてるわけですね。だからワークシェアリングこれはオランダなんかでもすでにもうだいぶ前からやっているわけですが、そういうことをやって解決していくことが健全ではないかと思ひます。

で AI における課題というのがございますね。これは皆さんが相当論じてこられたので、やめておきますけども。ロボットが責任主体になりうるかという問題。それから昔は天の神様が人たちの行動を全てみてて、けしからんことをやってもちゃんと罰するということを天の神様がやってた。ところが今はロボットとかネットワーク、情報システムがあらゆる人のあらゆる行動を見張ってる、そういう時代になってきたわけでありまして、どなたかの先生がデジタル神様とか書いておられましたけども、そういう問題をこれからどう考えるかという深刻な問題がございます。でそれがプライバシーの問題

題にもかかわる。

それからもう一つ大事な問題は、バーチャルリアリティー、それからゲームマシンのようなものですね、或る意味で麻薬的な効果をもつわけですね。だからこれをどういう風に規制するかというのはこれから非常に深刻な問題になる。つまり小学校の生徒とか中学校の生徒がゲームに夢中になって学校なんか行かないで部屋にこもって一日中やっている、そしたらちょっと頭が狂ってくるというようなこともあり得るわけなんで。そういうことが社会の現象としてはっきり出てきている今日、麻薬的な側面をもっているそういうコンピューターの世界バーチャルリアリティーの世界そういうものを本当にどうしていったらいいかっていうことに関して、もっと世論を戦わせて考えていかなければならない。これはヨーロッパとかユネスコでもそうでしょうけど、或る程度議論が始まっているという風に聞いておりますけれども、日本でもこういう問題を深刻にとらえて議論する必要があるんじゃないかという風に思います。

それからもう一つだけ。AI に対して義務とか制約を課すということを考える必要が出てきているんじゃないかということですね。つまり人間の尊厳とか人間を支配する可能性を排除する。で人間が危険と感じることをさせないとか人の嫌がる仕事とか過激な労働を AI に任せるとか。そんなのは当然ですけれども、AI が社会的な格差の助長につながっていかないように考える。特にこれからはっきりさせていかなければいけないのは、国際的な紛争とか戦争にロボットが使われる問題です。これは事実そろそろ始まっているわけですが、これは本当に国際世論によってやらせないようにしなければ、無茶苦茶なことになっていく可能性があるということを十分認識する必要があるということです。場合によってはそういったことをさせないという内容をロボット、AI システムの中に事前に装備していくということを義務づけるぐらいのところまでもっていかざるを得ないんじゃないか。昨日でしたかのお話でデザインということを考えなければいけないというお話がございましたけども、ロボットとかAI のデザインというときに、色んなデザインの仕方があるけども、その中にやっちゃいけないことを義務づけるということをデザインの中に入れるということを考えないといけないんじゃないかということでございます。

AIは哲学の最前線
AI as a Forefront of Philosophy

長尾 真
京都大学名誉教授

哲学はこれまで何を論じてきたか

- 宇宙の成り立ち
- 人間とは何か ——→ 人工知能
- 存在とは何か

哲学は頭脳の根源的な働きに向かって進んできた

- 大脳皮質の働き（論理の世界）
- 古代ギリシャからカント、ヘーゲルまで
- 感性、感情の世界（心の世界）
- キル・ケゴールからメルロ・ポンティまで
- 無意識の世界
- フロイドからピアジェまで
- 言語哲学、分析哲学
- 心理学、認知科学から魂の世界へ

哲学のための手段の展開

- 対話を通じて議論を深める 古代ギリシャ哲学
- 書物を通じて思考を深める 近代哲学、言語哲学
- 実験を通じて人の心を知る 心理学、認知科学
- 論理学などを用いて議論に客観性を持たせる 言語哲学、分析哲学
- コンピュータを使ってモデルを作る 人工知能

AIは人間頭脳の働きについての実証的哲学

- 現在のAIは大脳皮質の働きを模擬することに中心がある（認識、推論、知識、学習）
- 学習のために大量のデータが必要
- 学習機能によってAIは飛躍的に進歩している
- AIは目的を与えられれば学習によって人間以上にそれを達成する可能性がある
- 将来は論理の世界から、感情など心の世界の解明とその利用の世界に移って行く

AIは人の心を理解できるか

- 感性、感情の検出はある程度できる
- それには、言語の理解、人の発話に含まれる意味、感情などが把握できねばならない
- 人の発話に適切に応答するAIを作ることが重要な課題である
- そのためには人の心の状態の適切な推定が必要である
- 介護ロボットは人の心を和ませる応答が出来ねばならない

AIは人の仕事を奪うか

- AIに任せられる仕事は人の心に関係しない仕事である
- AIに任せられる仕事は創造的でない仕事である
- AIの導入によって得られる利益の一定の割合を税としてとり、失業者に対するベーシック・インカムの財源としてはどうか
- それよりもワーク・シェアリング制度を作り、それに参加する人たちにベーシック・インカムを分け与えるほうが良いだろう

AIにおける課題

- ロボットは責任主体になりうるか
- 現代は神に代わってロボットやネットワークが人間のあらゆる行為を見守っていることの問題
- プライバシーの問題
- VRなどは人の心を捉え、現実社会から遊離させる麻薬的側面を持つ
- 人の健全な心とは何かをよく考え、情報社会の種々の環境が人の心に与える影響について深刻に検討すること

AIに対して義務・制約を課すこと

- 人間性の尊重、人間を支配する可能性の排除
- 人が危険と感じることをさせない
- 人の嫌がる仕事、過酷な労働をAIによって代替え
- AIが社会的格差助長の原因にならないように配慮
- 紛争、戦争への利用についての禁止
- これらを社会のコンセンサスにする
- これらの中でプログラム化できるものはAIシステムに装備することを義務にする