対立を鎮めていった末に、軍事支出はほぼ不要となったし、限られた資源をめぐる競争に端を発する旧来の対立の原因は、繁栄によってほぼすべて解消された。この新世界秩序に取り込まれるのを拒んで激しく抵抗する独裁者も何人かいたが、いずれも、入念に組織された軍隊や大衆の反乱によって転覆させられた。

オメガズは、地球の生命史上もっとも劇的な変化を成し遂げた。史上初めてこの惑星は、たったひとつの権力に司られるようになったのだ。それに力を与える知能はあまりにも強力で、地球上や宇宙全体で何十億年にもわたって生命を花開かせる能力を備えていた。しかし、具体的にはどのような計画を持っていたのだろうか？

• • •

以上がオメガチームの物語である。本書ではこれ以降もうひとつの物語を示していくが、その物語はまだ紡がれていない。それは、AIとともに生きる我々自身の未来の物語だ。あなたはどんな展開を期待するだろうか？　オメガズの物語に少しでも似たようなことが、実際に起こりえるのだろうか？　あなたはそれを望むだろうか？　超人的なAIをめぐる憶測は別として、どんな物語の出だしが好ましいだろうか？　今後数十年で、AIが労働や法律や軍事にどんな影響をおよぼすのがよいのだろうか？　さらに先を見通して、どんな結末を書くことになるのだろうか？　これはまさに宇宙の調和に関する物語で、ほかならぬこの宇宙における生命の究極の未来がかかっている。その物語は我々が紡いでいくのだ。

第**1**章

いまもっとも重要な議論へのいざない

テクノロジーは生命に、かつてなく繁栄する力、
または自滅する力を与える。
――生命の未来研究所（FLI）

この宇宙は誕生から１３８億年経って目覚め、自己を認識するようになった。この宇宙の中で意識を持ったちっぽけな一部分が、小さな青い惑星から望遠鏡を使って宇宙のあちこちを見つめはじめ、それまで万物と考えていたものが実はもっと壮大な存在の一角にすぎないことを次々に明らかにしていった。太陽系、銀河、そして、数千億もの銀河が銀河群や銀河団や超銀河団といった複雑なパターンで連なった宇宙。自己意識を持つ彼ら天体観測者たちは、多くの事柄をめぐって互いに意見を異にしながらも、これらの銀河が美しくて荘厳であるという点ではおおむね一致している。

しかし、美というのは物理法則でなく見る人の目に基づいているので、この宇宙が目覚めるまで美は存在していなかった。それだけに、この宇宙が目覚めたことはいっそう不思議だし、喜びに値する。この宇宙を、自己意識がなく心を持たないゾンビから、自己認識や美や希望を備え、目標や意義や目的を追求する生態系へと一変させたのだから。もしこの宇宙が目覚めていなかったら、宇宙は完全に無意味で、空間の莫大な無駄遣いだっただろうと思う。仮に宇宙規模の大災害か、または自らが招いた災難によって、この宇宙が永遠の眠りに逆戻りしたら、悲しいことに再び宇宙は無意味になってしまうだろう。

しかしその一方で、ますます発展する可能性もある。我々人類がこの宇宙で唯一の天体観測者なのか、さらには最初の天体観測者なのかはまだ分からないが、この宇宙に関するこれまでの知見から考えるに、宇宙はこれまでよりもはるかにはっきりと目覚める可能性を秘めている。現在の我々は、今朝あなたが目を覚ましたときに感じた微かな自己意識のようなものだ。それから目を開いて完全に目覚め、もっとはっきりした意識を持つ。もしかしたら生命はこの宇宙全体に広がって、何十億年も、何兆年も繁栄するかもしれない。そしてもしかしたら、この小さな惑星上で我々が生きているうちに下す決定が、それをもたらすのかもしれない。

複雑さのおおまかな歴史

では、この驚異の目覚めはどのようにして起こったのだろうか？　それは単発的な出来事ではない。この宇宙を次々に複雑で興味深いものに仕立て上げていった、１３８億年におよぶ絶え間ないプロセス、いまも加速度的に続いているプロセスの、１ステップにすぎないのだ。

物理学者である私は、幸運にも過去四半世紀のほとんどを費やしてこの宇宙の歴史の解明に力を尽くしてきた。それは驚きの発見の旅路であった。私が大学院生の頃には、この宇宙の年齢は１００億歳なのか２００億歳なのかという程度の議論だったが、そこから望遠鏡とコンピュータと知識の発展のおかげで、１３７億歳なのか１３８億歳なのかと論じられる段階にまで進んできた。ビッグバンが何によって引き起こされたのか、そしてそれが真の万物の始まりだったのか、あるいはそれ以前の段階の結果だったのか、我々物理学者はまだ断定できていない。しかしビッグバン以降の出来事については、度重なる高精度の測定によってかなり詳細に解明されているので、ここで少々時間をもらって、

138億年におよぶ宇宙の歴史を簡単にまとめさせてほしい。

はじめに光があった。ビッグバンから1秒も経っていない時期、我々が望遠鏡で原理的に観測できる空間領域全体（「観測可能な宇宙」あるいは単に「この宇宙」と呼ぶ）は、現在の太陽の中心部よりもはるかに高温で明るく輝き、急激に膨張していた。壮観だったように思えるかもしれないが、素粒子が完全に均一に混じった、生命などどこにもいない高温高密度のスープにすぎなかったという意味では、退屈な状態でもあった。どこもほぼ同じ様子に見え、興味深い構造といったら、ランダムに見える微かな音波がところどころのスープの密度を約0・001パーセント高くしているくらいだった。その微かな音波はいわゆる量子ゆらぎとして生じたと広く考えられている。量子力学におけるハイゼンベルクの不確定性原理によると、完全に退屈で均一な状態というものは許されないのだ。

宇宙が膨張して冷えるにつれ、素粒子が組みあわさって次々に複雑な物体となり、宇宙はどんどん興味深い場所になっていった。最初の1秒足らずのあいだに、強い核力によってクォークが結合して陽子（水素の原子核）と中性子になり、その一部が数分のうちにさらに結合してヘリウムの原子核となった。およそ40万年後、電磁気力によってその原子核と電子が結合して最初の原子ができた。宇宙が膨張しつづけるにつれて、それらの原子は徐々に冷えて暗く冷たいガスとなり、その第一夜の暗闇はおよそ1億年続いた。この長い夜が明けたのは、重力によってガスのゆらぎが増幅され、原子どうしが引きあって最初の恒星や銀河が作られたときだった。それらのファーストスターは、水素の核融合によって炭素や酸素やケイ素といったもっと重い原子を作りながら、熱と光を発生させた。それらの恒星が死ぬと、使われていた原子の多くは再利用され、第2世代の恒星のまわりをめぐる惑星を形作った。

ある時点で原子の一群が複雑なパターンに配列し、自らを維持して複製できるようになった。すぐにコピーがふたつでき、次々に2倍ずつ数を増やしていった。わずか40回の複製で1兆個に達するため、この最初の自己複製体はあっという間に見過ごせない力を持った。こうして生命が誕生した。

生命の3つの段階

生命をどのように定義するかという疑問は、激しい論争を呼ぶことで悪名高い。相異なる定義がいくつもあるし、細胞から構成されているといったきわめて具体的な条件を含む定義は、未来の知能マシンや地球外文明には当てはまらないかもしれない。本書では、生命の未来をめぐる考察を、我々がこれまでに出合ったことのあるような生物種に限定させたくはない。そこで代わりに、単に「自身の複雑さを維持して複製できるプロセス」と、きわめて幅広い形で生命を定義しておこう。複製されるのは物質（原子からできている）ではなく、原子の配置を規定する情報（ビットからできている）である。細菌が自らのDNAのコピーを作るときには、新たな原子が作られるのではなく、新たな一群の原子がオリジナルと同じパターンに並ぶことで、情報がコピーされる。つまり生命は、情報（ソフトウェア）によってその振る舞いとハードウェアの設計図が決定される、自己複製する情報処理システムととらえることができる。

この宇宙そのものと同様、生命も徐々に複雑で興味深いものに変わっていった。* そしていまから説明するとおり、生命は洗練度に応じて、ライフ1・0、ライフ2・0、ライフ3・0という3つのレベルに分類すると都合が良い。図1・1にその3つのレベルをまとめてある。

この宇宙でいつどこでどのようにして最初の生命が誕生したかはいまだ明らかでないが、地球上の生命はおよそ40億年前に出現したことを示す強力な証拠がある。それからまもなくして、地球には多種多様な生命形態があふれかえった。その中で成功したものがすぐにほかを圧倒し、何らかの方法で環境に対応できるようになった。具体的に言うとそれらの生命形態は、感覚器で外界の情報を集め、その情報を処理して、どのように環境に反応するかを決定する、コンピュータ科

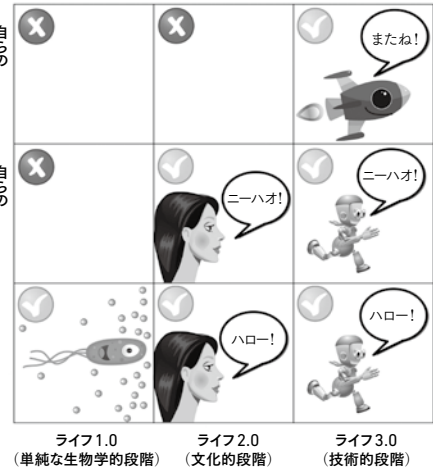

図1.1 生命の3つの段階：生物学的段階、文化的段階、技術的段階。ライフ1.0は、生きているうちに自らのハードウェアもソフトウェアも設計しなおすことはできない。どちらもDNAによって決まっており、何世代にもわたる進化によって変化するのみである。それとは対照的にライフ2.0は、自らのソフトウェアの大部分を設計しなおすことができる。人間はたとえば言語やスポーツや職業など新しい複雑な技能を習得できるし、自らの世界観や目標を根本から改めることもできる。地球上にはまだ存在していないライフ3.0は、自らのソフトウェアだけでなくハードウェアも大幅に設計しなおすことができ、何世代もかけて徐々に進化するのを待つ必要はない。

学者が「知的エージェント」と呼ぶものである。その情報処理の中には、あなたが目と耳から得た情報を使って、会話の中で何をしゃべるかを決定するといった、かなり複雑なものも含まれる。その一方で、きわめて単純なハードウェアやソフトウェアしか必要としないものもある。

たとえば多くの細菌は、周囲の液体中の糖濃度を測定する感覚器を持っており、鞭毛（べんもう）と呼ばれるプロペラ型の構造体を使って泳ぐことができる。その感覚器と鞭毛とをつなぐハードウェアは、「糖濃度感覚器が数秒前よりも低い値を報告してきたら、鞭毛の回転を反転させて方向を変えよ」といった、単純だが有用なアルゴリズムを実装したものかもしれない。

あなたは発話など数え切れない技能を習得しているが、それに対して細菌はたいして学習できない。細菌のDNAは、糖の感覚器や鞭毛などのハードウェアのデザインだけでなく、ソフトウェアのデザインまで規定している。細菌が糖に向かって泳ぐのは、けっして習得したからではなく、最初からDNAにそのアルゴリズムがコードされているからだ。もちろん何らかの学習プロセスは存在したが、ある1匹の細菌が生きているうちに起こったわけではない。何世代にもわたるゆっくりとした試行錯誤のプロセスを経て、糖の摂取量を高めるようなランダムなDNA変異が自然選択によって選

＊＝なぜ生命は次々に複雑になっていったのか？　進化は、環境の規則性を予測して利用できるような複雑さを持つ生命に恩恵を与えるため、環境が複雑であればあるほど、複雑で知的な生命が進化する。その賢くなった生命はさらに複雑な環境を作り、競合しあう生命形態がさらに複雑に進化することで、やがてきわめて複雑な生態系が形作られるのだ。

ばれることで、その細菌の種が進化する、その過程によって起こったのだ。変異の中には、鞭毛などハードウェアのデザインの改良に寄与するものもあったし、糖を見つけるアルゴリズムなどのソフトウェアを実装した情報処理システムを改良させるものもあった。

このような細菌に代表されるのが、私が「ライフ1・0」と呼ぶ生命、すなわち「ハードウェアとソフトウェアの両方が、設計されるのではなく進化する生命」である。それに対してあなたや私は、「ライフ2・0」、すなわち「ハードウェアは進化するだけだが、ソフトウェアの大部分は設計できる生命」である。あなたのソフトウェアとは、感覚から得た情報を処理してどんな行動を取るかを決定するのに使われる、アルゴリズムと知識全般のことだ。つまり、友人を見てそれを友人と認識することから、歩いたり読んだり書いたり、計算したり歌ったりジョークを言ったりすることまで、ありとあらゆる能力のことである。

あなたは生まれたときにはこのどの課題もこなせなかったのだから、そのソフトウェアはすべて、我々が学習と呼んでいるプロセスを通じて、のちに脳にプログラムされたものである。子供の頃には、あなたが何を学ぶべきかを決める家族や教師によって学習課程がおおむね設定されるが、その後、自分のソフトウェアをデザインする力を徐々に獲得していく。あなたの学校では学ぶ外国語を選択できただろう。フランス語を話せるようになるソフトウェアモジュールを自分の脳にインストールしたいか、それともスペイン語にしたいか？　テニスを習いたいか、それともチェスを習いたいか？　シェフになるための勉強をしたいか、それとも弁護士か、あるいは薬剤師か？　ＡＩや生命の未来のことをもっと知るために、それに関する本を読みたいか？

ライフ2・0は自らのソフトウェアをデザインする能力を持っているため、ライフ1・0よりもはるかに賢くなれる。高い知能を持つには、大量のハードウェア（原子からできている）と大量のソフトウェア（ビットからできている）の両方が必要である。我々人間のハードウェアのほとんどは生まれたあとに（成長によって）付け加えられるので、最終的な身体の大きさが母親の産道の幅によって制限されることはない。それと同様に、我々人間のソフトウェアのほとんどは生まれたあとに（学習によって）付け加えられるので、最終的な知能は、受胎のときに1・0スタイルのＤＮＡを介して受け継がれる情報の量によって制限は受けない。私のいまの体重は生まれたときの約25倍だし、脳の中のニューロンどうしをつなぐシナプス結合は、生まれたときに受け継いだＤＮＡの約10万倍の情報を保存することができる。あなたのシナプスが情報量にしておよそ１００テラバイト相当の知識や技能をすべて保存しているのに対し、ＤＮＡが保存できる情報量は約1ギガバイト、ダウンロードした映画1本がかろうじて収まる程度だ。そのため、乳児が生まれながらにして完璧な英語をしゃべったり、大学の入学試験でトップの成績を取ったりするのは物理的に不可能である。両親からもらったメインの情報モジュール（ＤＮＡ）に十分な情報保存容量がないので、その情報を脳にあらかじめロードしておく術はないのだ。

ライフ2・0は、自らのソフトウェアをデザインできるおかげで、ライフ1・0よりも賢くなれるだけでなく、柔軟にもなれる。環境が変化したら、ライフ1・0は何世代もかけて徐々に進化して適応するしかないが、ライフ2・0はソフトウェアのアップデートによってほぼ瞬時に適応できる。たとえば、抗生物質とたびたび出くわす細菌は何世代もかけて薬剤耐性を進化させることができるが、1個の細菌が行動を変えることはけっしてない。それに対して、自分はピーナッツアレルギーだと

知った少女は、瞬時に行動を変えてピーナッツを避けるようになる。この柔軟性のおかげでライフ2・0は、集団レベルでますます優位に立つ。人間のDNAに保存された情報は過去5万年でさほど劇的には進化していないが、集団として脳や書物やコンピュータに保存された情報は爆発的に増えてきた。高度な音声言語を介したコミュニケーションを可能にするソフトウェアモジュールをインストールしたことで、ある人の脳に保存されているきわめて有用な情報をほかの脳にコピーして、もとの脳が死んでからも残せるようになった。読み書きを可能にするソフトウェアモジュールをインストールしたことで、記憶しておけるよりもはるかに多くの情報を保存して共有できるようにもなった。そして、（たとえば科学や工学を学んで）テクノロジーを生み出すための脳のソフトウェアを編み出したことで、世界中の多くの人間が、何回かクリックするだけで世界中のほとんどの情報にアクセスできるようになった。

この柔軟性のおかげで、ライフ2・0は地球を席捲することができたのだ。遺伝的な足枷（あしかせ）から解放された人類全体の知識は加速度的なペースで増えつづけ、言語、文字、印刷機、現代科学、コンピュータ、インターネットなど、ひとつひとつのブレイクスルーが次なるブレイクスルーを可能にしてきた。このように我々の共有するソフトウェアが文化として加速度的に進化することが、我々人類の未来を決める支配的な力となったために、遅々とした生物学的進化はほぼ取るに足らないものとなったのだ。

しかし、人類は今日きわめて強力なテクノロジーを持っているが、我々が知るすべての生命形態はいまだに生物学的なハードウェアの基本的制約を受けている。100万年生きつづけたり、ウィキペディアの内容をすべて記憶したり、既知の科学を残らず理解したり、宇宙船に乗らずに宇宙飛行を楽しんだりできる人なんて一人もいない。生命がほぼ存在していないこの宇宙を、何十億年も何兆年も

繁栄する多様な生物圏に変え、この宇宙の潜在力を発揮させて完全に目覚めさせることのできる人などいない。このいずれを実現させるにも、生命が最後のアップグレードをしてライフ3・0になり、自らのソフトウェアだけでなくハードウェアもデザインできるようになる必要がある。つまりライフ3・0は、自身の運命を司って、進化の足枷からようやく完全に解放される存在なのだ。

生命の3つの段階を隔てる境界線は少々ぼやけている。細菌がライフ1・0で人間がライフ2・0であるのなら、ネズミはライフ1・1に分類できるかもしれない。ネズミはいろいろな事柄を学ぶことができるが、言語を編み出したりインターネットを発明したりできるほどではない。しかも言語を持っていないので、せっかく学んだことも死んでしまえばほとんど失われて、次の世代に受け継がれることはない。同様に、現代の人間はライフ2・1ととらえるべきだと言う人もいるかもしれない。人工の歯や膝やペースメーカーを埋め込むなどしてハードウェアを少しだけアップグレードすることはできるが、身長を10倍に伸ばしたり脳の大きさを1000倍にしたりするといった劇的なアップグレードはけっしてできない。

まとめると、生命が自らをデザインする能力に応じて、生命の進化は以下の3つの段階に分けることができる。

ライフ1・0（生物学的段階）――ハードウェアとソフトウェアが進化する。
ライフ2・0（文化的段階）――ハードウェアは進化するが、ソフトウェアの大部分はデザインされる。
ライフ3・0（技術的段階）――ハードウェアとソフトウェアがデザインされる。

138億年にわたって宇宙が進化してきた末に、ここ地球上でその進歩が劇的に加速している。ライフ1・0は約40億年前に、ライフ2・0（我々人間）は約10万年前に登場し、多くのAI研究者が考えるところでは、ライフ3・0は次の世紀、もしかしたら我々が生きているあいだにも、AIの進歩によって誕生するかもしれない。どんなことが起こり、我々にとってどういう意味があるのか？それが本書のテーマである。

論争

ライフ3・0は何をもたらすか？――この疑問は激しい論争の的（まと）になっており、世界を代表するAI研究者たちの見解は、未来予測だけでなく感情的な反応についても、自信たっぷりの楽観から深刻な不安に至るまで、互いに激しく食い違っている。AIが経済や法律や軍事にどのような影響をおよぼすかといった短期的な疑問についても、見解の一致はみられていない。もっと未来に目を向けて、「汎用人工知能（AGI）」、とくに、人間のレベルに到達して人間を凌ぎ、ライフ3・0を生み出すようなAGIをめぐる疑問となると、意見の不一致はますます広がる。「汎用知能」とは、たとえばチェスを打つプログラムといった狭い（特化型）知能とは対照的に、ほぼあらゆる目標を達成できるものを指す。

おもしろいことに、ライフ3・0をめぐる論争は、ひとつでなくふたつの別々の疑問を中心に渦巻いている。「いつ？」と「何？」、つまり「ライフ3・0は（もし出現するとしたら）いつ出現するか？」と「人類にとってどういう意味を持つのか？」という疑問である。私が見たところ、世界を代表する専門家が大勢属していて真剣に耳を傾けなければならない学派が3つある。図1・2に示したとおり、「デジタルユートピア論者」「技術懐疑論者」「有益AI運動の活動家」である。いまからそれぞれの主要な代弁者を紹介していこう。

デジタルユートピア論者

私は子供の頃、億万長者というのは威張っていて傲慢なものだと思っていた。しかし2008年にグーグルのラリー・ペイジと初めて会って、そんな先入観はすっかり打ち砕かれた。ジーンズとごくありふれたシャツというカジュアルな恰好で、ラリーはMIT（マサチューセッツ工科大学）の持ち寄りパーティにすっかり馴染んでいた。その思いやりのある穏やかな口調と親しげな笑顔を見て、おびえるどころか気楽に話ができた。2015年7月18日、イーロン・マスクとその当時の妻タルラーがカリフォルニア州のナパ・バレーで開いたパーティの席でラリーと偶然再会したときには、子供がうんちに興味を持っているという話になった。私がアンディ・グリフィスの大名著『ぼくのおしりがイカれた日（*The Day My Butt Went Psycho*）』を薦めると、ラリーはその場で早速注文した。ラリーが史上もっとも影響力のある人物として歴史に名を残すかもしれないことなど、私はついつい忘れてしまった。超知能デジタル生命が私の生きているうちにこの宇宙を支配するかどうか、それはラリーの決断如何だと思う。

ラリーと私は、それぞれの妻ルーシーとメイアと一緒に夕食を終えてから、機械はいずれ必然的に意識を持つようになるのかを議論しあったが、ラリーによればその疑問はまやかしだという。夜も更けてカクテルを何杯もやったあと、ラリーとイーロンのあいだでは、ＡＩの未来について、そして何をすべきかについて活発な議論が延々と続いた。時計の針が12時を回ると、見物人や茶々を入れる人がどんどん増えていった。ラリーは、私が言うところの「デジタルユートピア論者」の立場を熱く擁護した。デジタル生命は宇宙の進化における次のステップとして自然で望ましいものであり、デジタルの心を抑圧したり奴隷にしたりするのでなく、解放してやれば、ほぼ間違いなく良い結果が訪れるという立場だ。

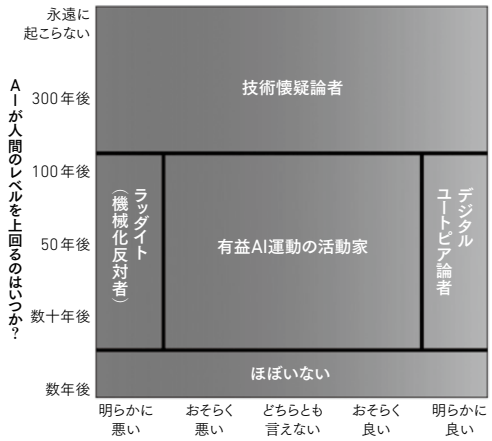

図 1.2　強い AI（あらゆる認知的課題で人間に匹敵する）をめぐる論争のほとんどは、「（出現するとしたら）いつ出現するか？」と「人類にとって良いことか？」というふたつの疑問を中心に渦巻いている。技術懐疑論者とデジタルユートピア論者は、心配する必要はないという点では一致しているが、その理由はまったく異なる。技術懐疑論者は、近い未来に人間レベルの AGI が出現することはないと確信しているが、デジタルユートピア論者は、AGI は出現するが、それはほぼ間違いなく良いことだと考えている。有益 AI 運動の活動家は、懸念を抱くことが未来のためになり、いまから AI の安全性に関する研究を進めれば良い結果になる確率が高まると感じている。ラッダイト（機械化反対者）は、悪い結果になると確信していて AI に反対している。この図はティム・アーバンからヒントを得た[1]。

ラリーはデジタルユートピア論者の中でももっとも影響力のある人物だと思う。生命は銀河系やその先まで広がっていくべきだと考え、そのためにはデジタルの形態でなければならないと訴えていた。ラリーがなによりも心配していたのは、ＡＩに対する被害妄想のせいでデジタルユートピアの実現が遅れたり、「邪悪になるな（Don't Be Evil）」というグーグルのかつてのスローガンに反するようなＡＩが軍事組織に独占されたりすることだった。それに対してイーロンは反論を続け、デジタル生命が我々の大切にしているものを破壊しないとそこまで確信している根拠を含め、主張の細部をもっとはっきりと示せと迫った。ラリーの方も何度かイーロンのことを、炭素でなくシリコンでできているというだけで劣った生命形態とみなす「種偏見論者」と非難した。この興味深い問題と議論については、第４章以降で詳しく掘り下げていこう。

その暖かい夏の晩のプールサイドではラリーのほうが分が悪かったようだが、ラリーが雄弁に擁護するデジタルユートピア論には大勢の有名な支持者がいる。ロボット研究者で未来学者のハンス・モラヴェックが１９８８年に書いた代表作『電脳生物たち』が、あらゆる世代のデジタルユートピア論者を奮い立たせ、その伝統は発明家のレイ・カーツワイルに受け継がれてさらに磨き上げられた。強化学習というＡＩの一分野の開拓者であるリチャード・サットンも、このあと紹介するプエルトリコでの会議で、デジタルユートピア論を熱烈に擁護した。

技術懐疑論者

もうひとつの重要な思索家グループもＡＩに懸念を抱いてはいないが、その理由はまったく違う。

超人的なＡＧＩを作るのはあまりにも難しくて、今後何百年も実現しないのだから、いま心配するのはばかげているという考えだ。私が「技術懐疑論」と呼んでいるこの立場をアンドリュー・エンは、「殺人ロボットの出現を怖がるのは、火星が人口過密になるのを心配するようなものだ」と見事に表現している。アンドリューは当時、中国版グーグル、バイドゥの主任研究者で、先日ボストンでの会議で話をしたときにも同じ主張を繰り返した。また、ＡＩのリスクに対する懸念は、ＡＩの進歩を遅らせる邪魔物になりかねないと思うとも語った。同様の意見は、ほかの技術懐疑論者、たとえば、ロボット掃除機ルンバや産業用ロボット「バクスター」の立役者である元ＭＩＴ教授ロドニー・ブルックスも表明している。面白いことに、デジタルユートピア論者と技術懐疑論者は、ＡＩのことを心配すべきではないという意見では共通していながら、それ以外に一致する点はほとんどない。ユートピア論者のほとんどは、人間レベルのＡＧＩは今後20年から１００年のうちに実現するかもしれないと考えているが、技術懐疑論者は、それは無知に基づく非現実的な夢だと斬って捨て、ＡＩが人知のおよばないレベルに進化するシンギュラリティ（技術的特異点）の予言を「おたくの携挙」と呼んでばかにすることも多い。２０１４年12月にある人の誕生パーティで会ったとき、ブルックスは、私が生きているうちにそれが実現することは１００パーセントありえないと言ってきた。後日、私がＥメールで「99パーセントじゃないと言い切れるのかい？」と問いただすと、「99パーセントどころじゃない。１００パーセントだ。絶対に実現しない」という返事が来た。

有益ＡＩ運動の活動家

２０１４年６月にパリのカフェで初めて会ったとき、スチュワート・ラッセルはまさに絵に描いたようなイギリス紳士だと感じた。能弁で思慮深くて穏やかな話し方だが、瞳の輝きには大胆さがにじみ出ていて、ジュール・ヴェルヌの１８７３年の名作『八十日間世界一周』に登場する私の子供時代のヒーロー、フィリアス・フォッグが現代に甦ったかのように思えた。存命中のもっとも有名なＡＩ研究者の一人で、この分野に関する標準的な教科書を共同執筆しているが、その慎み深さと優しさにはすぐに心を許せた。スチュワートは、ＡＩの進歩のスピードを考えると今世紀中に人間レベルのＡＧＩが出現する可能性は間違いなくあり、期待は抱いているものの良い結果になるという保証はないと話してくれた。何よりも先に答えを出さなければならない重要な問題がいくつかあるが、きわめて難しい問題なので、必要となるまでに答えが得られるよう、いまから研究を始めるべきだという。

現在ではスチュワートのこの考え方が比較的主流で、世界中のいくつものグループが、スチュワートの説くＡＩ安全性研究を進めている。しかし以前からそうだったわけではない。ワシントン・ポスト紙のある記事によると、ＡＩ安全性研究が主流になったのは２０１５年からだという。それ以前は、ＡＩのリスクについて語ろうものなら、主流のＡＩ研究者からは誤解され、不安を煽ってＡＩの進歩を邪魔しようとするラッダイト（機械化反対者）だとして無視されていた。第５章で掘り下げるが、スチュワートと同様の懸念は半世紀以上前、コンピュータの先駆者アラン・チューリングおよび、チューリングと一緒に第２次世界大戦中にナチスドイツの暗号の解読に取り組んだ数学者のアーヴィ

ング・J・グッドによって、初めて論じられた。ここ10年間は、この手の問題に関する研究はおもに、エリエゼル・ユドカウスキーやマイケル・ヴァッサーやニック・ボストロムなど、本職のAI研究者でない一握りの思索家によって進められてきた。しかし主流のAI研究者の大部分は、彼らの研究からほとんど影響を受けなかった。さらに知能の高いAIシステムを作るという日々の課題にばかり集中して、成功した際の長期的な影響についてはじっくり考えなかったのだ。何らかの懸念を抱いていた人も知りあいのAI研究者の中にはいたが、その多くは、人騒がせなテクノロジー恐怖症と受け取られるのを恐れて声を上げようとはしなかった。

私は、このような分断した状況を変えてAIコミュニティ全体が団結し、有益なAIをいかにして実現させるか、その議論を主導する必要があると感じた。幸いにも、そう考えていたのは私一人ではなかった。2014年春に私は、妻のメイア、友人の物理学者アンソニー・アギーレ、ハーバード大学の大学院生ヴィクトリヤ・クラコフナ、Skypeの創業者の一人ヤーン・タリンとともに、生命の未来研究所（Future of Life Institute――以降FLIとする）という非営利団体を立ち上げた。目標は単純。生命の未来が続いて、できるだけ素晴らしいものになるようにすることである。具体的に言うと、テクノロジーの力によって生命はかつてないほど繁栄するか、または自滅するかのどちらかだと感じ、我々は前者を望んだ。

第1回の会合は2014年3月15日、ボストン地区の学生や大学教授や思索家およそ30人が私の家に集まってブレーンストーミングをした。そして、バイオテクノロジーや核兵器や気候変動にも関心を払うべきだが、最初の大きな目標はAI安全性研究を主流にすることであるという点で、おおまかな合意が得られた。クォークの振る舞いの解明に貢献してノーベル賞を受賞した、MITの物理学者で私の同僚のフランク・ウィルチェックは、この問題に人々の関心を向けさせて無視されないようにするために、まずは新聞に署名入りの論説を書いたらどうかと提案してきた。そこでスチュワート・ラッセル（このときはまだ会ったことがなかった）と、私の仲間の物理学者スティーヴン・ホーキングに声をかけると、二人とも、私とフランクと並んで共同筆者になろうと請けあってくれた。それから推敲を重ねたが、ニューヨーク・タイムズなどアメリカの何紙もの新聞に掲載を拒否されたため、私がハフィントン・ポスト紙に持っているブログアカウントに投稿した。すると嬉しいことに、創業者のアリアナ・ハフィントン本人がEメールで、「震え上がりました！　トップに掲載しましょう！」と言ってくれた。こうしてフロントページのいちばん上に掲載されたことを皮切りに、その年の末にかけてマスコミがAIの安全性について次々と報じ、イーロン・マスクやビル・ゲイツなどテクノロジー界のリーダーたちも賛同してくれた。その年の秋にニック・ボストロムの著書『スーパーインテリジェンス』が出版されたことも、人々の議論の高まりをさらに後押しした。

FLIの有益AIキャンペーンにおける次の目標は、世界中の代表的なAI研究者を会議に招くことで、誤解を払拭して合意を形成し、建設的なプランを組み立てることだった。そのような著名な人々に、会ったこともない外部の人間が主催する会議に来てくれと説得するのは容易ではないだろうと分かっていたし、異論のある議題だけになおさらだったため、我々はできる限りの策を講じた。マスコミの参加を禁じ、開催地を2015年1月のビーチリゾート（プエルトリコ）に設定し、参加費を無料にし（太っ腹なヤーン・タリンのおかげだ）、考えつく限りなるべく人騒がせでない題目として「AI

の未来――機会と困難」というタイトルを付けた。そして何よりも、タッグを組んでくれたスチュワート・ラッセルの働きで、学界と産業界両方のＡＩ指導者が何人も組織委員会に加わってくれて、規模を大きくすることができた。その一人、グーグルの「ディープマインド」（ＡＩ開発に携わるグーグルの系列企業）に所属するデミス・ハサビスはのちに、囲碁でもＡＩが人間に勝てることを見せつける。デミスのことを知れば知るほど、彼はＡＩを強力にするだけでなく、有益なものにするという野心を抱いているのだということがはっきり分かってきた。

こうして、錚々たる面々が集まる素晴らしい会議が実現した（図1・3）。ＡＩ研究者が、一流の経済学者や法学者、テクノロジー界のリーダー（イーロン・マスクなど）、およびさまざまな思索家（第4章で取り上げる「シンギュラリティ」という言葉を作ったヴァーナー・ヴィンジなど）と顔をあわせたのだ。そして、我々のもっとも楽観的な予想をも上回る成果が得られた。陽光とワインのおかげだったのかもしれないし、タイミングがちょうど良かったのかもしれない。異論の多いテーマでありながら驚くほど意見が一致し、それをまとめた公開書簡[2]には最終的に８０００人を超える人々が署名して、さながらＡＩ界の人名録のようになった。書簡の主意は、ＡＩ研究の目標を定めなおして、方向性のないＡＩでなく、有益なＡＩを作ることを目標とすべし、というものである。書簡にはまた、会議参加者が合意した、目標達成のための研究テーマの詳細なリストも挙げた。こうして、有益ＡＩ運動は主流になった。その後の進展については本書のあとのほうで紹介しよう。

この会議で学べたもうひとつの重要なこと、それは、ＡＩ研究の進展によって浮かび上がってきた数々の疑問は単なる知的興味の対象ではなく、我々の選択が生命の未来全体に影響をおよぼしかねない、倫理的にもきわめて重要な事柄であるということだ。人類が過去に下してきた選択の倫理的影響力は、確かに大きいものもあったが、すべて限りがあって、我々は最悪の疫病からでさえ立ち直ったし、史上最大の帝国でさえやがては崩壊した。太陽が明日も昇るのと同じように、明日の人類も、貧困や病気や戦争といった絶え間ない災難に直面しながらも立ち直ることを、過去の世代は知っていた。しかしプエルトリコの会議で講演し

図1.3　2015年1月にプエルトリコで開催した会議には、AIやその関連分野の一流研究者が集結した。後列左から、トム・ミッチェル、シーン・オヘイガルテー、ヒュー・プライス、シャミル・シャンダリア、ヤーン・タリン、スチュワート・ラッセル、ビル・ヒバード、ブレイス・アグエラ・イ・アルカス、アンダース・サンドバーグ、ダニエル・デューイ、スチュワート・アームストロング、ルーク・ミュールホイザー、トム・ディーテリッヒ、マイケル・オズボーン、ジェイムズ・マニカ、アジェイ・アグラワル、リチャード・マラー、ナンシー・チャン、マシュー・プットマン。後列以外の立っている人たち、左から、マリリン・トンプソン、リチャード・サットン、アレックス・ウィスナー＝グロス、サム・テラー、トビー・オード、ヨッシャ・バッハ、カティヤ・グレース、エイドリアン・ウェラー、ヘザー・ロフ＝パーキンス、ディリープ・ジョージ、シェーン・レッグ、デミス・ハサビス、ヴェンデル・ヴァラッハ、チャリーナ・チョーイ、イリヤ・サツケヴァ、ケント・ウォーカー、セシリア・ティリ、ニック・ボストロム、エリック・ブリニョルフソン、スティーヴ・クロッサン、ムスタファ・スレイマン、スコット・フェニックス、ニール・ヤコブシュタイン、マレー・シャナハン、ロビン・ハンソン、フランチェスカ・ロッシ、ネイト・ソアレス、イーロン・マスク、アンドリュー・マカフィー、バート・セルマン、ミシェル・ライリー、アーロン・ヴァンデヴェンダー、マックス・テグマーク、マーガレット・ボーデン、ジョシュア・グリーン、ポール・クリスティアーノ、エリエゼル・ユドカウスキー、ディヴィッド・パークス、ローラン・オルソー、J.B.ストローベル、ジェイムズ・ムーア、ショーン・レガシック、メイソン・ハートマン、ハウイー・レンペル、ディヴィッド・ヴラデック、ヤコブ・シュタインハート、マイケル・ヴァッサー、ライアン・カロ、スーザン・ヤング、オワイン・エヴァンズ、リヴァ＝メリッサ・テズ、ヤーノシュ・クラマー、ジェフ・アンダーズ、ヴァーナー・ヴィンジ、アンソニー・アギーレ。しゃがんでいる人たち、サム・ハリス、トマソ・ポッジョ、マリン・ソリャシウ、ヴィクトリヤ・クラコフナ、メイア・チタ＝テグマーク。撮影：アンソニー・アギーレ（隣にしゃがんでいる人間レベルの知能のそばにフォトショップで合成した）。