호모 파베르(Homo Faber)와 호모 에티쿠스(Homo Ethicus) 사이에서*

-알파고(AlphaGo)에 대한 대중의 두려움을 중심으로

김진형**

주제분류 기술철학, 공학윤리

주 요 어 알파고, 인공지능, 행동주의, 인간 존엄성, 호모 파베르, 호모 에티쿠스, 호모 사피엔스

요 약 문

2016년 3월에 벌어진, 알파고와 이세돌의 바둑 대결이 있은 후, 국내의 주된 반응은 인공지능에 대한 두려움이었다. 이 논문에서 필자는 그 두려움의 정체를 간단히 분석한 후, 그것을 토대로 세 가지를 주장한다.

첫째, 알파고(인공지능)에 대한 대중의 두려움은 공연하다.

둘째, 우리는 알파고를 인간처럼 대해야 한다.

셋째, 인공지능에 의한 인류의 종말이 현실화된다면, 그것은 호모 사피엔스로서의 인간의 숙명이다.

* 이 논문을 꼼꼼히 읽고, 생산적인 지적과 의문으로 논지를 좀 더 정교하게 하는 데 도움을 주신 익명의 심사위원들께 감사드린다. 아울러 일부 오해를 야기한 부분은 전적으로 필자의 책임이라는 점을 밝힌다.

** 서울시립대

1.

2016년 3월 9일부터 5일간, 프로기사 이세돌 9단과 구글 딥마인드(Google Deep Mind)가 개발한 인공지능프로그램 알파고[1]가 대국을 벌였다. 이때 특히 국내에서의 반응이 어떠했는지는 언론 매체의 기사 제목들인 다음에서 잘 알 수 있다.

> "소름끼치는 직관과 추론, 우린 알파고의 영혼을 보았다"(정재승, <한겨레> 2016.03.12.)
>
> "인공지능은 인류를 지배할 것인가"(조홍섭, <한겨레> 2016.03.12.)
>
> "인공지능 '치명적 한계' 셋"(권오성, <한겨레> 2016.03.15.)
>
> "알파고 충격? 진짜 걱정은 따로 있다"(<주간동아> 2016.03.16. 1029호)
>
> "인공지능은 축복일까, 재앙일까" (박정경, <중앙일보> 2016.03.25)

이 반응들은, 기계가 정작 기계 같지 않다는 놀라움으로 시작하여, 모종의 두려움을 거쳐 그 두려움을 순화하려는 자구적인 노력, 즉 반은 감정적이고 반은 이성적인 노력으로 이어진 후, 마침내는 사태의 본질을 차분히 살피는 이성적 성찰로 정리될 수 있다.

그런데, 알파고와 이세돌의 대국을 '기계 vs. 인류'의 대결 구도[2]로 보

1) 알파고는 1200여대의 중앙처리장치(CPU)가 연결된 수퍼 컴퓨터로, 빅데이터 연산을 수행하는 소프트웨어다. 알파고는 초당 경우의 수 10만개를 검색할 수 있는 것으로 알려져 있다.

2) 이러한 반응이 얼마나 현실적인지는 관련 교양도서가 곧장 출간되어 인기를 얻은 데서 알 수 있다. 대표적으로는 다음을 들 수 있다. 김대식, 『김대식의 인간 vs

고 이세돌의 패배를 가슴 아파하며 TV 앞에서 승리를 간절히 응원하던 대중들의 모습과 전문 바둑 기사들의 다소 편파적인(?) 해설을 고려할 때, 위의 반응들을 관통하는 핵심은 바로 '두려움'이다. 물론 그것은 다음과 같은 즉각적인 인식에서 기인하는 두려움이다.

> 이세돌보다 뛰어난 능력을 지닌 알파고가 사람이 아닌 기계이다.

2.

대중들의 두려움은 알파고가 '사람'이라면, 예컨대 중국의 커제(柯潔·19) 9단이었다면 갖지 않았을 두려움이다. 그렇다면 왜 '알파고'이기 때문에 두려워하는가? 혹은 왜 '기계'이기 때문에 두려워하는가?

쉽게 알 수 있듯이 이 두려움은 이중적이다.

> 두려움1: 알파고가 인간보다 '뛰어난' 능력을 지녔다는 사실에 대한 두려움
> 두려움2: 뛰어난 능력을 지닌 알파고가 인간이 '아니라는'(즉 기계라는) 사실에 대한 두려움

먼저 두려움1을 보자. 그 원인으로 가장 먼저 생각해볼 수 있는 것은 기계와 인간의 전쟁 시나리오이다. 달리 말하면, 영화 [터미네이터]에서와 같은 인간에 대한 인공지능 기계의 공격을 상상하기 때문이다. 인간보다 뛰어난 능력을 지닌 존재가 있다면, 그것이 무엇이든 약자로서의 인간은 공격당할 가능성, 따라서 멸종의 가능성을 늘 열어둘 수밖에 없을 것이다. 그렇다면 기계는 정말로 인간을 공격할까?

기계, 인공지능이란 무엇인가』, 동아시아, 2016.

이 시나리오가 현실이 되기 위해서는 적어도 다음이 성립해야 한다.

> 인류가 알파고(인공지능)의 존속과 번영에 위협을 가하거나 알파고가 인류를 방해물로 간주한다.

그러나 이것이 참일 가능성은 없어 보인다. 무엇보다 인간은, 자신보다 뛰어난 신체기능과 지적능력을 가진 것이 분명한 존재, 따라서 승산이 희박한 싸움을 걸 정도로 어리석지 않기 때문이다. 더 나아가 인간은 스스로가 도덕적 행위의 주체임을 천명하는 만큼, 그리고 적어도 그 천명을 스스로 뒤집지 않고 있는 만큼, 공연한 이유로 기계의 존속과 번영에 '위협'을 가하지 않을 것이기 때문이다. 다음으로 인간은 추론을 통한 뛰어난 예측 능력 역시 가지고 있으므로 무고한 인간을 방해물로 간주하는 비도덕적인 기계를 만들만큼 어리석지도, 그것을 통제하지 못할 만큼 지적으로 무능하지도 않을 것이기 때문이다. 즉, 도덕적인 기계를 만들 수 있기 때문이다. 이 점은 실제 문제의식과 함께 그에 따른 노력도 꾸준히 이어지고 있다는 데에서도 알 수 있다. 흔히 선도적인 예로 꼽히는 것이 바로 잘 알려진 '로봇3원칙'[3]이다.[4] 결론적으로, 도덕의식과 뛰어난 지적 능력을 바탕으로 한 이러한 적극적인 노력을 고려할 때, 두려움1은 허상이라고 할 수 있다. 그럼에도 불구하고 두려움1에서 자유롭지 못하다면, 그것은 곧 인류는 똑똑하지 않거나 도덕적이지 않다는 것, 그리고 앞으로도 거기에 머물고 만다는 것을 스스로 인정하는 것이다.

---

3) 3원칙은 다음과 같다. <1.로봇은 인간에 해를 가하거나, 행동을 하지 않음으로써 인간에게 해가 가도록 해서는 안 된다. 2.로봇은 인간의 명령에 복종해야한다. 다만, 그 명령들이 첫 번째 법칙에 위배될 때에는 예외로 한다. 3.로봇은 자신의 존재를 보호해야한다. 다만, 그러한 보호가 첫 번째와 두 번째 법칙에 위배될 때에는 예외로 한다.> 이 원칙의 문제점이 대두된 이후에 상위 원칙으로서 다음이 추가되었다. <0. 로봇은 '인류'에게 해를 가하거나, 행동을 하지 않음으로써 인류에게 해가 가도록 해서는 안 된다.>

4) 국내 관련 연구로는 다음을 참고하기 바람. 고인석(2014), 변순용(외)(2017).

물론 기계, 즉 인공지능은 전쟁이 아닌 다른 방식으로 인간에게 위협적인 존재일 수 있다. 그것은 바로 지금까지 우리가 맡아왔던 일자리의 대체이고 따라서 일자리의 상실이다.[5] 그러나 이 또한 과도한 상상이다. 왜냐하면 인류의 존속과 번영은 인간 능력을 효율적으로 대체하는 도구의 발견 내지 발명과 그것의 발전, 그리고 대안적 일자리의 창출로 가능하였으며, 지금까지도 그 역사는 이어지고 있기 때문이다. 이를 뒷받침하는 전망으로는 예컨대 다음을 들 수 있다.[6]

> 1, 2차 혁명이 생존 욕구를 위한 물질혁명이고, 3차 혁명이 관계 욕구를 위한 인터넷 연결 혁명이라면 (인공지능에 의한) 4차 혁명은 (자아표현과 자아실현의 욕구로서의) 경험 욕구를 위한 정신 소비 혁명이 될 것이다. 기술의 진보에 대항하는 일자리는 사라진 반면, 기술의 진보는 새로운 시장을 창출함으로써 일자리 또한 만들어왔다. 그런데 4차 산업혁명은 정신 소비의 시장을 창출할 것이다. 따라서 4차 산업혁명은 새로운 일자리를 창출할 것이다.

3.

그렇다면, 뛰어난 능력을 지닌 알파고가 사람이 아니라는 사실에 대한 두려움2는 어떤가? 우선 드는 의문은, 도대체 기계를 두려워한다는 것이 말이 되는가이다. 예를 들어, 우리는 인간의 한계를 뛰어 넘는 기능을 가진 비행기를 두려워하거나 단시간에 최적인 길을 찾아 순간순간 목적지

5) 예컨대, 2016년 다보스 포럼에서 710만 개의 일자리가 선진국에서 사라질 것이라 예측하였으며. 옥스포드 대학은 미국 일자리의 47%가 20년 내 사라질 것이라 예측하고 있다. 이민화(2016) 14쪽 참조.

6) 이민화(2016), 17~19쪽 참조. 이 논증에서 논란거리는 마지막 전제라고 할 수 있다. 필자는 정신의 향유를 지향하는 인간의 특성이 고려된 그 전제에 대한 이민화의 논의에 동의한다.

를 안내해주는 앱(App)이 장착된 스마트폰을 두려워하지 않는다. 말하자면, 기계 자체는 결코 두려움의 대상이 아니다. 물론 핵무기 시스템과 같은 기계는 두려움을 낳기도 한다. 하지만 그것은 사용에 관련된 두려움이지 기계 자체, 즉 도구 혹은 수단에 대한 두려움이 아니다. 그런데 사용에 관한 두려움이라면, 우리는 사실상 인류가 만들었거나 만들고 있는 모든 도구를 두려워해야 한다. 가령, 플라스틱조차도 부적절한 사용과 처리로 인해 인류에게 위협이 될 수 있기 때문이다.[7)]

이 점을 고려할 때, 두려움2는 다음과 같이 재 서술될 수 있다.

> 두려움2*: 인간이 아닌 알파고가 인간처럼 행동한다는 사실에 대한 두려움
>
> 실제로 바둑 해설가들은 알파고를 사람처럼 대했다.
>
> "알파고가 하상귀는 포기하는 것처럼 보이네요"
> "이번 교전에서는 만족스럽다고 판단한 걸까요, 우상귀로 옮겨가네요."
> "이세돌의 가운데 대마를 견제하기 위해 여기에 포석을 둔 것 같네요."

여기에서 언급되고 있는 '행동(behavior)'은, 예컨대 '바둑판의 특정한 위치에 바둑돌을 가져다 둠'이라고 할 수 있다. 그런데 그 행동이 단순한

7) 이근영, "지금까지 플라스틱 총생산량 83억톤 대부분 쓰레기로 버려져", <한겨레>, 2017. 7. 21. 이 비유와 관련 한 익명의 심사위원은 원전이나 핵의 파괴력의 크기를 들어 플라스틱과의 비교가 적절하지 않다고 말하였다. 하지만 이 지적은 잘못이다. 가령, 플랑크톤에 축적되고 있는 미세플라스틱의 위험성은, 비록 순간적으로 현실화되지는 않지만 환경론자들의 우려처럼 인류를 위협할 정도로 크기 때문이다. 참고로, 위험 분석 전문가들에 따르면 호랑이보다는 모기가 더 위험하다. 해리스 외(2006), 7장 참조. 한편, 위험에 대한 전문가들의 인식과 일반인들의 인식의 차이에 관한 국내의 흥미로운 연구로는 차용진(2007)을 참조하기 바람.

움직임을 뜻한다면 그것은 두려움을 야기할만한 것이 아니다. 왜냐하면, 예컨대 반려견에 의해서도 그러한 단순한 움직임은 얼마든지 이루어질 수 있기 때문이다. 따라서 알파고의 행동은 모종의 '생각'과 관련된 동작으로 이해해야 한다. 이 경우 두려움2의 또 다른 버전은 다음과 같다.

두려움2#: 인간이 아닌 알파고가 인간처럼 생각한다는 사실에 대한 두려움

두려움2*와 두려움2#의 구분에서 두려움2의 정체를 알 수 있다. 즉, 두려움2는 기계인 알파고에게서 관찰되는 특정한 행동(B1)에, 인간에게만 귀속된다고 여겨지는 특정한 생각(M1), 즉 일종의 심성(mentality)이 대응된다는 판단에서 오는 것이라고 할 수 있다. 물론 그러한 판단은 매우 그럴듯하다. 왜냐하면 그것은, 대중들 스스로가 의도했든 의도치 않았든, 심리철학에서의 행동주의(behaviorism)[8]로 설명될 수 있기 때문이다. 행동이 심성을 구성한다고 주장하는 그 이론에 따를 때, 대중들의 판단은 다음과 같은 논변에 기인한다.

인간의 행동 B는 심성 M1을 구성한다. 그런데 알파고의 행동 B1은 인간의 행동 B와 다르지 않다. 따라서 B1은 M1을 구성한다.

이제 만일 누군가(편의상 이세돌이라고 하자)가 알파고의 행동을 관찰함으로써 두려움2를 느낀다고 하자. 그러면 그에게 알파고는 부지불식간에 인간과 다르지 않은 존재로 다가온 것이다. 알파고는 인간의 '그' 생각(M1)을 하고 있기 때문이다.[9] 다시 말하면, 이세돌이 두려움2를 가지

8) 행동주의에 따르면, 어떤 마음 혹은 심성을 갖는다는 것은 적합한 패턴의 관찰 가능한 행동을 표출하거나 표출하는 성향을 갖는 것의 문제이다. 김재권(1999), 52쪽 참조.

9) 알파고의 행동B1과 심성 M1의 대응에 관한 판단이 정당하다는 것은 해설가들의

고 있다는 것은 그가 알파고를 인간과 다르지 않은 존재로 인지한다는 것을 뜻한다. 따라서 두려움2는 사실상 다음과 같다.

기계가 인간일 수 있다는 사실에 대한 두려움

실제로 많은 사람들은 이러한 두려움을 가지고 있다. 그렇다면, 기계가 인간일 수 있다는 것(혹은 기계가 인간과 다르지 않다는 것)이 정말로 두려워할 만한 사실인가? 다시 말하면 이세돌은 두려움2를 가질 이유가 있는가? 일단, 답은 "없다"이다. 왜냐하면 인간이 아닌 것으로 생각했던 것이 인간과 다르지 않을 수 있다면 그것은 놀랍거나 신기한 일이지 두려워할 일은 아니기 때문이다.

하지만 여전히, 알파고가 두려움의 대상일 여지는 있다.

4.

단지 기계로서든 아니면 인간과 다르지 않을 수 있는 존재로서든, 알파고가 두려움 내지 우려를 낳는 다른 까닭은 무엇일까? 무엇보다, 알파고에 의한 '인간 존엄성'[10]의 침해 가능성을 꼽을 수 있다. 그런데 과연

표현인 "이세돌이 우상귀에 돌을 놓았네요."에 대한 "이세돌이 우상귀가 현상황에서 최선의 착점이라는 생각을 하였다."라는 말의 관계에 의해서도 알 수 있다. 왜냐하면 후자는 이세돌의 심성에 관한 진술로서 청자로 하여금 이세돌이 현 시점에서 어떤 심적 상태에 있는지를 이해하도록 하는데, 이는 후자의 진술이 전자의 진술로 번역될 수 있다는 것을 뜻하기 때문이다. 이러한 번역 가능성은 철학자 헴펠(Hempel, C)의 주장으로서 그것은 다음과 같이 표현된다. <모든 심적 현상들을 서술하는 유의미한 심리 진술은 어떤 것이든지 오직 행동적인 현상과 물리적인 현상만을 서술하는 진술로 번역될 수 있다.> Hempel(1980)

10) 인간 존엄성의 귀속과 관련한 입장들로는 다음을 들 수 있다. "인류에 속하는 모든 구성원에게 인정된다.", "인격체에게만 인정되어야 한다."이에 관한 간략한 이

그러할까? 혹시 그러한 두려움이나 우려 또한 공허한 것이거나 적어도 착시의 결과는 아닐까?

먼저, '기계 vs. 인간'이라는 구도에 갇혀 있는 경우를 보자. 앞서 언급한 일자리 대체 결과로서의 특정 개인들의 실업 상태나 인간에 대한 기계의 공격을 인간의 존엄성 침해로 보는 것은 온당치 않아 보인다. 왜냐하면, 그러한 사태는 기계라는 바로 '그' 이유에서 특별히 발생하는 것이 아니기 때문이다. 즉 알파고에게만 귀속되는 책임으로서의 고유한 침해가 아니기 때문이다.

논의의 편의상, 최근까지도 논란거리가 된 '존엄사'에서 '존엄성' 문제를 생각해보자. 일단 존엄사는 '품위 있는 죽음'을 뜻한다. 구체적으로는 '인간으로서 갖추거나 갖추어야 할 품위를 손상하지 않는 죽음'을 뜻한다고 할 수 있다. 그러면 이때의 품위는 곧 존엄성의 한 예화이다. 이제 존엄사가 '소극적이고 자의적인 안락사'와 사실상 다르지 않다는 점을 고려해보자. 그러면 '품위'를 훼손하는 요소들로는 다음을 들 수 있다.

(ㄱ) 무의미한 생명연장
(ㄴ) 죽음을 선택할 권리의 제한
(ㄷ) 타인에 대한 피해(혹은 타인의 이익 침해)
(ㄹ) 기계 장치에 의존하는 무기력함
(ㅁ) 물리적•심적 고통

물론 더 많은 항목이 가능하다. 중요한 것은 알파고의 존재가 이러한 항목의 어느 것을 구체화하는가이다. 혹은 알파고가 '기계'라는 바로 '그' 이유로 인해 이 항목에서 읽을 수 있는 '품위'를 훼손하는가이다. 즉, 알파고로 인해 인간의 '품위'있는 죽음, 즉 '존엄사'는 금지되거나 불가능해지는가? 오히려, 잘 발달된 도구로서의 알파고라면, 존엄사가 필요한 특

해를 위해서는 다음을 참조하기 바람. 구인회(2002), 22-26쪽.

정 상황을 극히 최소화할 여지, 혹은 위의 요소들을 해소할 여지가 크다.[11] 그렇다면, 알파고가 인간과 다르지 않을 수 있다는 그 사실은 어떤가? 이 요소들을 함축하는가? 즉 품위를 훼손하는가? 쉽게 알 수 있듯이 그렇다고 보기 어렵다.

이번에는 보다 넓은 의미에서, 인간의 '자율성' 침해를 인간 존엄성 훼손으로 보기로 하자. 인간이 아니지만 인간과 다르다고 볼 수 없는 어떤 존재가 있다는 사실은 독립적이고 독자적인 존재로서의 인간의 '자율성'의 침해 내지 소멸을 함축하는가? 혹은 그것은 적어도 자율 혹은 자유의지의 제한을 함축하는가? 만일 그렇다고 해도 그 경우가 다른 존재에 의한 침해와 다르다고 볼 합당한 이유를 찾기는 어려워 보인다.[12] 따라서 인간 존엄성 훼손 측면에서의 두려움은 사실 공허하거나 오도된 것일 가능성이 크다.

그럼에도 불구하고 현실에서, 알파고는 대중들로 하여금 인간의 존엄성 문제를 상기시키고 있다. 물론 만약에 알파고가 이세돌에게 전패했다면, 존엄성 문제는 대두되지 않았을 것이다. 알파고는 단지 바둑 두기를 어리숙하게 흉내 내는 정도에 지나지 않는, 따라서 올바른 바둑 두기 능력을 갖추지 못한 '그저 그런' 존재일 뿐이기 때문이다. 다시 말해서 '그저 그런' 존재를 보면서 결코 '그저 그렇지 않은' 존재가 자신의 존엄성 문제를 떠올릴 리는 없기 때문이다.

그런데 이세돌은 사실상 전패했고, 그 즈음에 그가 한 말은 "인간이

11) 레이 커즈와일에 따르면, 기술의 기하급수적 진보로 인해 2030년이면 수명이 대폭 늘어난 인류(New Type)가 등장한다. 김정욱(외). "인공지능이 일을 하고 인간은 500살까지 생명 연장할 것", <매일경제>, 2017.07.08. (http://news.mk.co.kr/newsRead.php?year=2017&no=457310 , (검색일: 2017.07.20)

12) 잘 알려진 대로 칸트(Kant, I)는『도덕형이상의 기초』에서 '자율(autonomy)'을 인간 존엄성의 근거로 본다. 김광연(2016) 참조. 이 경우, 알파고가 인간 존엄성을 침해한다는 것은, 곧 '자율'을 박탈하거나 불가능하게 한다는 것으로 이해해야 한다. 그러나 필자의 입장은 과연 알파고의 존재가 어떤 점에서 그러한 일을 초래할지는 알기 어렵다는 것이다.

진 것이 아니라 이세돌이 진 것이다"였다. 그리고 커제 9단은 "프로 바둑기사들의 존엄을 위해 전력을 다해 일전을 치르겠다"라고 말했다. 먼저, 이세돌의 말에는 인간은 알파고가 넘을 수 없는 능력을 지닌 존재라는 암시가 있다. 커제의 말에는 당연히 그 능력의 구체적인 한 예가 언급되어 있다. 말하자면, 바둑 두기는 인간만이 가진, 고차원적이고 심오하며 신비롭기도 한 독특한 능력이 요구되는 행위라는 것이다. 이는 프로 바둑기사들의 존엄을 위하는 길이, 프로 바둑기사일수록 바로 '그' 능력을 더 잘 갖추고 있음을 확인시키는 일, 즉 압도적인 승리임을 뜻한다. 물론 가끔 있을 수 있는 인간의 패배는 단지 실수로 인한 우연한 결과가 된다. 만일 그 고유한 능력의 확인에 실패하거나 실패할 수밖에 없다면, 그들의 존엄은 유지될 수 없다.

이 점에 주목할 경우, 인간의 존엄성 문제가 발생하는 이유로는 다음을 생각할 수 있다.

> 인간이 아닌 존재가 인간만의 고유한 능력을 가졌거나 가질 수 있다는 사실.

이것은, 알파고가 인간이기 위한 조건을 갖춘다는 것, 따라서 사실 인간의 정체성 내지 유일성이 사라진다는 것을 뜻한다.[13] 이세돌과 커제의 말, 그리고 거기에 공감한 대중들은 바로 그것을 인간의 존엄성 훼손으로 여기는 것일 수 있다. 달리 말하면, 프로 바둑기사들을 쉽게 이기거나 적어도 대등한 수준의 알파고(또는 장래의 알파고)를 마주하는 순간, 인간은 그 독자적인 지위를 더 이상 고집할 수 없게 된다는 데에서 존엄성 훼손 더 나아가 상실을 느낄 가능성이 높다.

---

13) 물론 알파고가 인간의 그 고유한 능력을 갖출 수 없다는 주장이 가능하다. 이 경우라면, 인간의 유일성 문제는 발생하지 않을 것이다. 이와 관련해서는 본문 아래의 '직관'에 관한 논의를 참고하기 바람.

그런데, 그러한 이세돌과 대중은 사실상 인간의 종(種) 우월성 내지 인간 중심주의를 암암리에 전제하고 있다. 하지만 종 우월성이 정당화될지, 따라서 알파고의 바둑 두기 능력이 과연 '감히' 가져서는 안 될 그러한 것인지는 매우 의심스럽다.[14] 결론적으로 두려움2, 나아가 존엄성 훼손 측면에서의 두려움 또한 공연한 것이다.

존엄성 문제는 알파고가 인간과 다르지 않을 수 '있다'는 데에서 더욱 부각될 수 있다.하지만 방금 보았듯이, 인간 존엄성의 훼손 우려는 정당화되기 어려운, 말 그대로 기우일 수 있다. 만일 인간과 다르다면, 이미 언급했듯이 그 문제는 애초부터 발생하지 않으며 따라서 오도된 판단이나 두려움도 없을 것이다. 따라서 알파고의 존재는 인간 존엄성 문제와 사실상 무관하다고 할 수 있다. 그렇다면 남은 문제는 우리에게 알파고는 무엇이고 또한 무엇이어야 하는가이다.

5.

먼저, 알파고는 인간과 다르지 않을 가능성이 크다.[15] 물론, 그렇지 않을 가능성도 있다. 사실, 알파고의 심성 혹은 생각(M1)은 인간의 그것과 같을 수 없음을 보이는 일은 철학적으로 이미 제시된 바 있다. 행동주의

14) 이러한 의심은, 예컨대 동물을 차별적인 존재로 보아서는 안 된다는 싱어(Singer, P)의 입장과 유사한 맥락에서 제기될 수 있다. 또한, 백종현에 따르면, 자연인이 가령 기계와는 다른, 대체 불가능한 존재라는 사실은 인간이 존엄하다는 주장의 근거일 수 없다. 백종현(2015), 147쪽 참조.

15) 필자는 '도덕성' 혹은 칸트가 말하는 '자율'이 인공지능과 인간을 가르는 기준인지에 대해서는 다루지 않았다. 기본 입장은 인공지능 또한 도덕적 주체일 수 있다는 것이지만 그것을 쟁점으로 삼는 것은 이 글의 범위를 넘어선다. 참고로, 홀(Hall, J.S)에 따르면, 인공지능로봇은 지적인면과 도덕적인 측면에서 다르지 않을 수 있다. 이것과 다른 입장들에 대해서는 다음을 참조하기 바람. Veruggio. G and Operto, F(2006)

의 세련된 형태[16]로 알려진 기능주의(functionalism)에 대해, 서얼(Searl, J)이 제기한 반론이 그것이다. 그의 '중국어 방 논증'은 튜링(Turing, A)이 제안한 사고 실험, 즉 튜링 테스트에 대한 반론으로서, 그 요지는 다음과 같다.

> 아무리 복잡하고 '지능적이고' 세련된 것일지라도 입-출력 프로그램을 수행하는 기계는 (인간의) 심성과 동등할 수 없다. 왜냐하면, 중국어 방에 있는 사람은 중국어를 '이해'한 것이 아니기 때문이다.[17]

이 반론의 핵심은, 알파고의 승리 혹은 패배는 알파고 자신과 이세돌의 수(手)를 '이해'한 결과가 아니라는 것이다. 이것은 인간의 지능과 알파고의 지능의 차이를 '이해' 여부에 둠으로써 알파고가, 최소한 지능 면에서는, 인간처럼 보이기는 해도 결코 인간과 같을 수는 없다는 주장에 해당한다.

하지만, 이 반론은 유효하지 않다. 달리 말하면, 알파고는 여전히 인간과 다르지 않을 수 '있다'. 왜 그러한가? 적어도 다음 두 가지가 옹호될 수 있어야 하지만 정작 그렇다고 보기 어렵기 때문이다.

> (a) 제4국에서 이세돌로 하여금 신(神)의 한 수로 평가되었던 78번째 수를 두도록 한 '이해'와 제2국에서 알파고로 하여금 결정타였던 37번째 수를 두도록 한 '계산'은 다르다.
> (b) 알파고의 지능이 '이해'를 하는 것은 현실적으로 불가능하다.

(a)를 옹호하기 위해서는 알파고의 신경망 구조[18]의 작동, 혹은 그것을

16) 기능주의가 행동주의 세련된 형태라는 것은 그 둘 사이에 차이점이 있다는 것을 뜻하기도 한다. 하지만 그러한 차이점은 본문에서의 논의 초점을 고려할 때 큰 관련이 없다.

17) 자세한 소개와 논의는 김재권(1999)의 4장과 거기에 언급된 튜링과 서얼 등의 문헌을 참조하기 바람.

기반으로 한 딥러닝(deep learning)이 이세돌의 사고와 같을 수 없음을 보여야 한다.

그리고 (b)를 옹호하기 위해서는 인공지능의 '이해' 능력을 함축하는 전문가들의 다음 주장이 틀렸다는 것을 보일 수 있어야 한다.

> (기술적) 특이점(singularity)은 인간을 초월하는 인공 지능이 태어나는 시점에 벌어지며, 인간은 자신을 초월하는 인공 지능이 어떤 의도를 가질지 짐작할 수 없기 때문에 특이점 이후의 세계가 어떤 모습일지 상상하는 것은 불가능하다.[19]

그러나 둘 모두 성공하기 어렵다. 먼저, (a)를 옹호하는 거의 모든 주장들이 취하는 핵심 가정은, 그 '알고리즘'은 인간이 짜서 입력했다는 바로 그 이유에서 인간의 사고와 다르다는 것이다. 그런데 이것은 기계의 '계산'이니까 인간의 사고가 아니라는 주장의 다른 표현에 지나지 않는다. 현재 쟁점은 <'계산'이 정말로 인간의 사고(이해)와 다른가?>인 만큼, 그 가정에 의존하는 것은 곧 선결문제요구의 오류를 범하는 것이며 따라서 설득력이 없다. (b)는, 특이점의 도래가 더욱 강하게 예측되는 현 상황과는 달리 그 도래의 '불가능성'을 입증할 이론은 정작 제시되고 있지 않다는 점에서, 설득력이 없다.[20] 결론적으로, 서얼의 반론은 유효하지 않다.

---

18) 신경망 구조는 인간의 뉴론 체계를 모방한 것으로서 정책망(policy network)과 가치망(value network)로 이루어져 있으며 이들의 작동을 통해 최적의 한 수를 찾는 것으로 알려져 있다.

19) 기술적 특이점에 관한 이해는 다음을 참조하기 바람. 레이 커즈와일(2007). 이 인용은 유신(2014), 221-221쪽의 것임. 괄호는 필자의 추가.

20) 기술공학적 측면에서, 특이점의 도래는 시기에 대한 논란은 있으나 여전히 긍정적이다. 이러한 흐름을 소개하고, 철학적 측면에서의 회의적인 견해 그리고 향후 도래할 미래에 대한 공학적, 철학적 고민의 개요와 관련된 전문 정보 등은 다음에서 접할 수 있다. 곽노필, "특이점이 온다면?…준비는 돼 있는 걸까" <곽노필의 미래창>, 2017.09.17

물론, 다른 방식으로 서얼의 반론을 유지할 수도 있다. 즉 '이해'는 인간만이 가지고 있는, 따라서 알고리즘으로는 결코 구현할 수 없는 어떤 것으로서의 '직관'을 포함한다고 주장하는 것이다. 이것은 예컨대 이세돌의 그 '78번째 수'는 '직관'의 산물인 반면, 알파고의 그 '37번째 수'는 그렇지 않다고 주장하는 것과 같다.

이제 여기서 이세돌의 '그' 직관이 다음과 같은 능력이라는 점에 주목해보자.

> 우주 내의 원자 수 보다 많은, 따라서 무한에 가까운 경우의 수 가운데에서, 계산 없이 승부를 결정짓는 바로 '그' 수를 찾는 능력. (바둑 기사들이 말하는 소위 '감'이자 프로기사일수록 더욱 잘 갖춘 능력)

이것을 과연 진정한 의미의 '능력'이라고 볼 수 있는가? 혹은 그것은 유의미한 인식적 작동인가? 예컨대 원주율을 구할 때, 무한한 자릿수 중에서 아직 계산되지 않은 1억조 번째 자리의 숫자를 '직관'을 이용해서 특정할 수 있는가? 특정했다면, 그것은 무작위 선택과 다른가? 다르다면 어떤 점에서 그러한가? 만일 무작위 선택과 다르지 않다고 하자. 그러면, 그 숫자가 나중에 실제 값으로 밝혀진다고 할지라도, 그것은 결코 '능력'의 산물이라고 볼 수 없다.

이세돌의 78번째 수가 무한에 가까운 경우의 수를 모두 두어 본 경험의 오류 없는 결과가 아니라면, '창의적'이라고 불리는 그 수는 사실 우연의 산물이라고 보아야 한다. 만일 그것이 '능력'이라면, 알파고 역시 정책망과 가치망을 작동하지 않음으로써 그러한 우연을 얼마든지 수행할 수 있으며, 따라서 '직관'할 수 있다. 즉 알파고의 한때 '떡' 수였던 '37번째 수'는 그러한 의미에서의 직관의 산물이었을 수 있고 따라서 알파고 역시 '이해'를 한다. 반면에 만일 그것이 '능력'이 아니라면, 이세돌의

그 '직관' 혹은 '감'은 '이해'에 포함되지 않으며 따라서 알파고와 인간의 지능이 다르다는 근거가 될 수 없다. 결론적으로, 알파고는 인간과 다르지 않을 수 있다.[21] 물론 그렇다고 해도, 앞서 언급한대로 존엄성 문제는 불거질 이유가 없다.

6.

알파고가 3연승을 한 후, 딥 마인드의 CEO인 허사비스(Hassabis, D)는 그 자신도 놀랐음을 표현하면서 알파고의 한계를 궁금해 했던 것으로 알려졌다.[22] 그 놀라움에서 기술적 특이점의 도래를 떠올린다면 그것을 과연 망상이라고 치부할 수 있을지는 의문이다. 또한, 그러한 도래에도 직관과 추상은 인간의 고유한 능력이라는 식의 주장을 한다면, 그것은 지능 혹은 이해의 의미를 인간에 맞추어 정의하려는 시도로서 핵심을 벗어나는 주장이다. 알파고의 37번 째 수는 적어도 가까운 장래에 '이해'의 결과일 수 있고 따라서 알파고와 인간은 적어도 이성 혹은 지적 측면에서 구분할 수 없을 것으로 생각되기 때문이다.

---

21) 알파고 혹은 인공지능로봇이 인간과 다를 수밖에 없는 다른 이유로는 '감정'을 들 수 있다. 예컨대 천현득은 '감정'을 가진 로봇은 당분간은 실현될 것이라고 보지 않는다. 그러나 그의 입장이 그 실현 가능성을 부정하는 것은 아니다. 천현득(2017). 실제로도 전문가들은 그 실현에 긍정적이다. 이에 대해서는 김평수(2016)을 참조하기 바람. 한편, 여기에서 한 익명의 심사위원은 다음과 같이 말한다. "인간은 생명체이고, 알파고는 기계이다. 따라서 ... '알파고는 인간과 다르지 않을 수 있다'는 필자의 주장은 비약이다." 하지만 이 지적도 역시 잘못되었다. 필자의 그 주장이 "기계가 인간과 생물학적으로 다르지 않을 수 있다"는 것이 아니기 때문이다. 따라서 이 지적은 이 논문에서 다루는 대상이 생명체가 아닌 알파고라는 단순한 사실을 인지하지 못한 데에서 온 것이라고 할 수 있다. 또한 필자는 본문에서 그 주장을 이끌어내는 논변을 충분히 제시하였다.

22) http://www.hani.co.kr/arti/sports/baduk/734618.html (검색일: 2017.07.19)

사실, 인간의 두뇌 기능에 대한 정보는 여전히 불완전하다. 불완전한 정보가, 알파고의 지능이 두뇌 기능에 해당하는지를 가리는 기준일 수는 없다. 물론 불완전한 정보라고 해도 본질적인 차이를 알려주는 정보는 존재할 수 있다. 그리고 그것이 인간과 인간 '아닌 것'을 가르는 기준일 수 있다. 하지만, 그것은 무엇인가?

알파고가 인간처럼 생각한다는 사실은 결코 두려워할 일이 아니다. 그러나 종우월주의 내지 인간중심주의적 사고에서 벗어나지 않는다면, 우리는 늘 불안한 존재일 수밖에 없다. 만일 앞서 언급했던 대중들의 두려움이 알파고가 그저 도구에 지나지 않을 때에도 사라지지 않는다고 하자. 그러면 그것은 인간이 매우 부조리한 혹은 이성적이지 못하다는 것을 뜻할 수 있다. 왜냐하면, 그것이 도구 자체에 대한 두려움이라면 그것은 도구적 인간(호모 파베르)임을 두려워하는 것이기 때문이다. 반면 대중들의 반응이, 도구의 사용에 대한 두려움이라면 그것은 윤리적 인간(호모 에티쿠스)으로서 자격이 없거나 윤리적 인간임을 스스로, 애초에 부정하는 것이기 때문이다.

결론은 다음과 같다. 인공지능은 결코 동물 종으로서의 사람이 아니다. 그러나 인간이 부조리한 존재가 아니라면, 인공지능의 '진화'를 신인류의 '등장'으로 받아들여야 한다.[23] 즉, 미래의 알파고를 인간**처럼** 대우해야 한다. 그 알파고(AI)는 애초에 인류가 자신을 위해 탄생시켰고, 이미 강조했듯이 두려워할 이유가 없으며, 그리고 무엇보다도 이 논문에 반론을

23) 각주21에서 언급한 한 익명의 심사위원은, 이 지점에서 또 다시 다음과 같은 취지의 지적을 한다. "인공지능이 동물 종으로서의 사람이 아니라는 주장과 인공지능을 신인류의 등장으로 받아들여야 한다는 두 주장은 모순이거나 논리적 비약이다." 이것은 필자의 주장을 "동물 종이 아닌 것을 동물 종으로 받아들여야 한다"는 것으로 이해할 경우에만 할 수 있는 지적이다. 그러나 필자는 이 논문 어디에서도 그러한 주장을 하지 않았고, 따라서 잘못된 지적이다. 결론적으로, 이상의 두 가지 지적을 이유로 '알파고를 인간처럼 대우해야 한다'는 주장이 정당화될 수 없다고 한 해당 심사위원의 평가는 오해에서 비롯되었음을 밝혀둔다.

제기할 '그' 철학자와 구별되지 않을 것이기 때문이다. 그리고 만일 인간 존엄성 문제가 실제로 발생하고 훼손된다면, 그것은 호모 파베르로서의 인류가 겪어야 할 당연한 시련이다. 그리고 그 시련을 어떻게 감당하고 해소할 것인가는 또한 호모 에티쿠스로서 마땅히 떠맡아야 할 과제이다. 더 나아가 만에 하나, 업그레이드 '된' 혹은 업그레이드 '한' 알파고로 인해 인류의 종말이 도래한다면 그것은 뛰어난 이성으로 지구에서 유일하게 번영할 수 있었던, 호모 사피엔스의 숙명[24] 이다.

24) 참고로, 향후 등장할지모를 소위, 강한 인공지능으로서의 알파고가 비도덕적인 존재일 경우를 생각해보자. 인류는 그 단초를 알아채어 업그레이드를 미리 멈추어야 한다. 그렇지 못하다면, 무능하거나 최소한 지적으로 불완전한 것이고, 따라서 지적 호기심에서 자유롭지 못한 호모 사피엔스로서 겪어야 할 숙명이다. 만일 강한 인공지능이 도덕적이라면, 인류의 종말은 오로지 인간이 비도덕적일 경우일 때에 일어날 것이다. 그런데 그러한 인류라면 당연히 제거되어야 할 처지에 놓일 것이다.

참고문헌

고인석(2014),「로봇윤리의 기본 원칙:로봇 존재론으로부터」,『범한철학』 제75집, 범한철학회.

곽노필, “특이점이 온다면?…준비는 돼 있는 걸까”, <곽노필의 미래창>, 2017.09.17

구인회(2002),『생명윤리의 철학』, 철학과현실사.

김광연(2016),「칸트 윤리학에 나타난 인간 존엄성의 근거와 보편가능성으로서의 도덕법칙의 요청」,『철학연구』 제138집, 대한철학회.

김대식(2016),『김대식의 인간 vs 기계, 인공지능이란 무엇인가』, 동아시아.

김재권(1999),『심리철학』, 하종호・김선희 옮김, 철학과현실사.

김정욱(외), “인공지능이 일을 하고 인간은 500살까지 생명 연장할 것”, <매일경제>, 2017.07.08. (http://news.mk.co.kr/newsRead.php?year=2017&no=457310, (검색일: 2017.07.20)

김평수 (2016).「인간과 교감하는 감성로봇 관련 기술 및 개발 동향」.『정보와통신』, 33(8), 한국통신학회.

레이 커즈와일(2007),『기술이 인간을 초월하는 순간 특이점이 온다』, 김명남・장시형 옮김, 김영사.

백종현(2015),「인간개념의 혼란과 포스트휴머니즘 문제」,『철학사상』 58, 서울대학교철학사상연구회.

변순용(외)(2017),「로봇윤리헌장의 필요성과 내용에 대한 연구」,『윤리연구』 제112호, 한국윤리학회.

유신(2014),『인공지능은 뇌를 닮아 가는가』, 컬처룩.

이근영, “지금까지 플라스틱 총생산량 83억톤 대부분 쓰레기로 버려져”, <한겨레>, 2017. 07. 21.

이민화(2016),「인공지능과 일자리의 미래」,『국제노동브리프』, 한국노동연구원.

차용진(2007), 「위험인식과 위험분석의 정책적 함의: 수도권 일반주민을 중심으로」, 『한국정책학회보』 16권 1호.

천현득(2017), 「인공지능에서 인공 감정으로 －감정을 가진 기계는 실현가능한가?－」, 『철학』 131, 한국철학회.

해리스 외(2006), 『공학윤리』, 김유신 외 옮김, 북스힐.

Hempel, Carl G(1980), "The Logical Analysis of Psychology," reprinted in *Readings in Philosophy*, vol. 1. ed. Ned Block, Cambridge: Havard University Press.

Veruggio, G and, Operto, F(2006), 'Roboethics: a Bottom-up Interdisciplinary Discourse in the Field of Applied Ethics in Robotics', *International Review of Information Ethics* Vol. 6.

http://www.hani.co.kr/arti/sports/baduk/734618.html (검색일: 2017.07.19)

Between Homo Faber and Homo Ethicus:
With a Focus on People's Fear of AlphaGo

Kim, Jinhyeong (University of Seoul)

As Lee Se-dol was defeated by Google's artificial intelligence algorithm AlphaGo in a historic Go match on Mar.,2016, people began to have a fear of artificial intelligence. In this paper, I critically examine the fear that they have. And then, I want to argue the following three:

First, you don't logically justify the fear of AlphaGo(AI).

Second, You should treat Alphago like us.

Thirdly, if an AI system leads to the fall of the human race, you, as specimens of Homo Sapiens, should take it as fate.

Key words: AlphaGo, Artificial Intelligence, Behaviorism, Human Dignity, Homo Faber, Homo Ethicus, Homo Sapiens.

김진형 e-mail: musoyu108@hanmail.net

투 고 일	2017년 10월 16일
심 사 일	2017년 10월 30일
게재확정	2017년 11월 15일