



사상의 지평

지금은 뉴욕타임즈의 정보기술 전문 기자로 일하고 있는 케이드 메츠는 2016년 3월 월간지 <와이어드(WIRED)>에 이세돌과 알파고(AlphaGo)의 바둑대국을 관전하고 다음과 같이 시작하는 서울발 기사를 올렸다. “2국에서 구글이 만든 머신은 인간이 전혀 뒤 본 적이 없는 수를 두었다. 그것은 대단한 한 수였다. 전 세계인들이 지켜보는 가운데 이 한 수는 오늘날 인공지능의 엄청난 능력과 신비로운 재능을 완벽하게 보여주었다. (중략) 19 × 19의 선의 교차점으로 이루어진 반상의 우변에 둔 37수는 이세돌을 비롯한 바둑의 고수들을 당황하게 만들었다.” 구글의 자회사 딥마인드(Deep Mind)가 만든 알파고는 이 대국에서 4:1로 이세돌을 이겼다. 이세돌은 두 번째 대국을 마치고 나서 “상당히 놀랐다. 할 말이 없을 정도가 아닌가 싶다.”고 말했다. 그러나 당시 스탠포드 대학 인공지능연구소장으로 있던 페이-페이 리(Fei-Fei Li)는 “나는 전혀 놀라지 않는다”면서 “자동차가 가장 빨리 달리는 인간보다 빠르다고 해서 놀랄 일인가?”라고 반문했다. 인공지능이 인간보다 더 똑똑하다면 놀랄 일인가?

대국이 시작되기 하루 전 마이크로소프트의 공동창업자인 폴 앨런이 돈을 대 만든 비영리 연구기관인 앨런 인공지능 연구소는 미국 인공지능진흥협회(Association for the Advancement of Artificial Intelligence) 회원들을 대상으로 설문조사를 했다. 쉰다섯 명의 응답자들 가운데 65%가 알파고가 이길 것이라고 응답했다. 언론은 “귀추가 주목된다”고 했지만 결과는 이미 예정된 것이었다. 결과 자체는 놀랄 일이 아니지만 ‘알파고vs이세돌’의 두 번째 대국에서 알파고가 둔 37수는 경험하거나 예상하지 못한 ‘신의 한 수’였다. 뒤늦게 알파고의 37수를 상기하는 것은 그것이 인간의 인식능력과 지능으로 인간이 만든 프로그램 혹은 알고리즘을 이해하지 못하게 되는 지능폭발(Intelligence Explosion)의 특이점과도 같은 사건이었기 때문이다.

인류의 문명사는 여러 간난과 곡절에도 불구하고 자연과 우주를 이해하고 여기에 대응하여 삶의 조건을 개선하는 일관된 방향으로 진행된 과정이었다. 그 출발은 우리가 살고있는 세계를 인간의 지능, 즉 사고력으로 인식하는 것이다. 삼라만상을 신의 섭리와 조화로 이해한 신비와 초자연적 인식의 단계에서 시작하여 물리적 인과관계로 해석한 합리적 이성의 시대를 거쳐 이제는 모든 것을 0과 1의 조합이라는 절대적 공리(axiom)로 환원하고 이를 퍼즐로 맞추듯 재구성하는 디지털의 시대로 진입하였다. 디지털 시대는 인류 문명사 전체로 볼 때 춘각에 불과한 짧은 시간이다. 앨런 튜링이 초보적 메카트로닉스 인공지능을 만든지 80년, 그리고 IBM의 인공지능 딥 블루(Deep Blue)가 체스 게임에서 체스 세계 챔피언 가리 카스파로프를 이긴지 불과 20년 만에 인류 문명은 다시 인간의 지적 능력으로는 이해할 수 없는, 신비의 세계로 넘어가는 사상의 지평(event horizon)에 이르렀는지도 모른다.

피타고라스와 그 후예들

세계를 숫자로 환원하여 이해하려는 시도는 고대부터 이루어져 왔다. 피타고라스와 그를 따르는 고대 그리스 철학자들은 대부분 수학자들이었다. 윤회를 믿는 신비주의자인 그들은 역설적으로 삼라만상이 숫자로 이루어졌다고 생각했다. 아리스토텔레스는 <형이상학(Metaphysics)>에 “피타고라스와 그의 제자들은 수학을 주창하였을 뿐만 아니라 완성했고 수학적 원리가 만물의 원리라고 생각했다”고 기록했다. 만물의 근원이 물이라고 주장한 탈레스를 필두로 하는 이오니아 학파가 유물론적이고 과학적이었다면 피타고라스 학파는 유심론적이고 추상적이었다. 인간이 오감을 통해 인식하는 사물을 극단적으로 추상하여 이론의 여지가 없는 명징성에 도달하려고 했던 것이다.

인간이 겪는 모든 갈등과 분란은 인간의 인식이 절대적이지 않고 상대적이라는 사실과 인간의 생각을 표현하는 언어의 모호성에 있다. 나는 “절반밖에 없다”고 생각하는데 너는 “절반이나 남았다”고 말한다. 내가 말하는 절반은 결핍이며 네가 말하는 절반은 충분이다. 이처럼 우리의 인식은 인식의 주체가 처한 시공간의 조건에 따라 달라진다. 일관성이 없다고 탓할 일만은 아니다. 상

황이 바뀌면 생각이 달라지는 것을 어쩔 것인가? 출처가 확인된 것은 아니지만 현대 거시경제학의 토대를 마련했다고 평가받는 케인즈(John Maynard Keynes)는 고위 정부 관리가 주장에 일관성이 없다고 비판하자 이렇게 대응했다고 전해진다. “사태가 바뀌면 저는 생각이 달라집니다. 당신은 어떤가요(When event changes, I change my mind. What do you do sir?)” 그렇게 보면 ‘내로남불’은 인간 인식능력의 불완전성을 보여주는 극단적 사례일 수도 있다. 자신의 인식능력, 그리고 그에 따른 판단이 불완전하다는 것을 인정하지 않으면 독선과 전횡 때로는 폭력적 강압에 빠지고 결국 때늦은 후회를 피할 수 없다. 단언컨대 인간에 관한 한 전지전능과 무오류는 없다. 왜냐하면 우리가 경험하는 세계는 아날로그적인 세계, 즉 유사체(類似體), 혹은 의사체(疑似體)들로 이뤄진 세계이기 때문이다. ‘그런 것’이 아니라 ‘그럴듯한 것’의 세계이다.

그러나 디지털의 세계에서는 이런 아날로그의 세계가 액시옴(axiom), 즉 절대불변의 수학적 공리로 환원되고 재구성된다. 알고리즘은 그런 해체와 재구성의 도구이다. 광대무변의 우주와 존재의 본질을 이해하려는 천체물리학이나 입자물리학은 철저하게 현상을 수학적 공리로 환원하는 과정이다. 피타고라스는 더 이상 환원할 수 없는 단위로 우리가 인식하는 구체적인 세계와 그 속에 존재하는 만물을 추상함으로써 진리에 도달하려고 했고 플라톤의 이데아 역시 수학적 추상에서 크게 벗어나지 않는다. 사물의 질서를 수학적 공리로 환원하려는 노력이 오랫동안 집적되어 우리가 인과관계를 인식하기 어려운 사회현상이나 우리가 우연이라고 말하는 것조차도 확률론, 게임이론 등의 수학적 명제로 정리될 수 있었다.



주역 : 디지털적 세계관

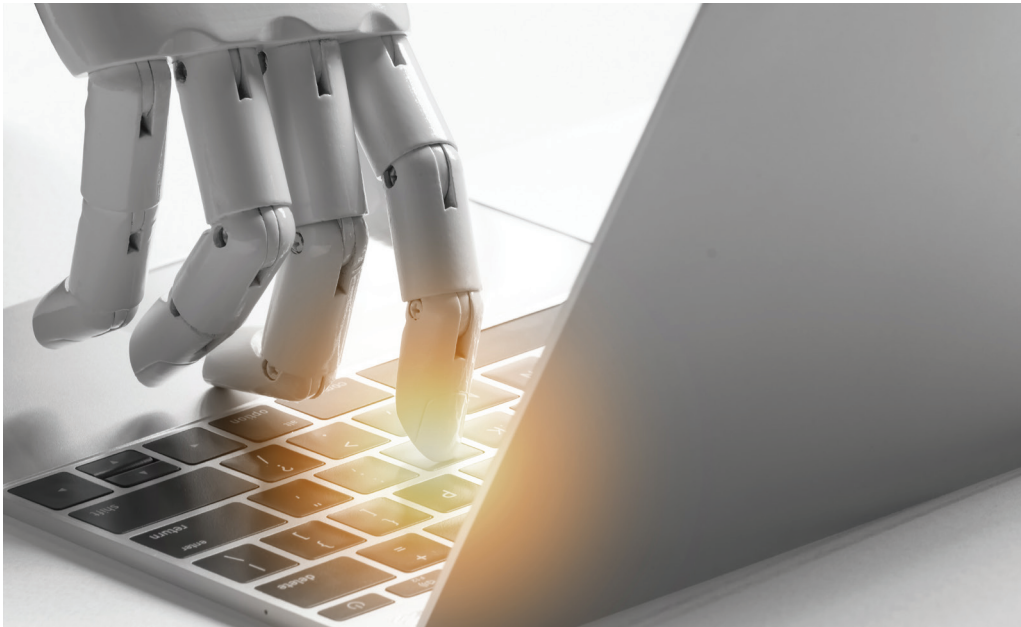
주역의 난해한 내용을 철학적으로 설명한 계사전(繫辭傳) 상편에 “역에는 태극이 있고 여기서 양의가 생긴다. 양의가 사상을 낳고 사상이 팔괘를 낳는다(易有太極, 是生兩儀, 兩儀生四象, 四象生八卦).”라는 말은 태극에서 비롯되어 음양이라는 둘의 조합으로 이뤄지는 세계관을 보여준다. 필자는 이것을 세계에 대한 디지털적인 해석이자 인간 언어의 한계를 극복하려는 노력에서 비롯된 것으로 본다. 말 자체도 모호성을 내포하고 있지만 이해타산에 따라 말이 달라지니 자연과 우주의 운행질서를 표현할 고정불변의 기호체계가 어찌 필요하지 않았을까.

고트프리트 라이프니츠 역시 논리, 즉 언어로 표기되는 인간의 사고체계를 절대적이고 명징한 기호체계로 치환하려고 했다. 그것을 보편부호(characteristica universalis) 혹은 보편언어라고 불렀다. 그는 1714년 1월 지인에게 보낸 한 서한에서 이렇게 썼다. “일종의 보편언어나 문자는 지금까지의 모든 언어와 전혀 다를 것이다. 왜냐하면 보편언어에서는 상징 그리고 심지어는 어휘가 합리성을 지향하기 때문이다. 사실의 오류를 제외하면 모든 오류란 단순히 계산상의 착오가 될 것이다. 이러한 언어 혹은 문자를 만들거나 고안하는 것은 매우 어렵겠지만, 사전이 없어도 매우 쉽게 이해할 수 있을 것이다.”

그는 베이징에 파견된 예수회 선교사들과 교류했다. 당시 북경에 체류하던 선교사 중에는 프랑스 황제 루이 14세가 보낸 조아킴 부브(Joachim Bouvet)가 있었다. 그는 당시 베이징에서 강희제에게 수학과 천문학을 가르치는 한편으로 역경을 연구했다. 역경을 연구하는 과정에서 그는 중국 송나라 때 역경에서 수리철학으로 발전시킨 사오용(邵雍)의 저서 <황극경세서(皇極經世書)>를 접했다. 그는 중국의 수리 철학에 대해 깊이 공감하고 라이프니츠에게 주역의 64괘에 대한 지식을 전해준다. 라이프니츠는 괄의 조합을 보고 0과 1의 조합으로 어떤 수든지 표기할 수 있다는 이진법에 대한 확신을 갖게 되었다. 그는 신이 창조한 세계를 이진법의 보편언어로 설명할 수 있다고 생각했다. 신이 만든 세계를 이진법적인 수리로 이해하려는 라이프치히의 시도는 자연의 선택을 통해 오늘날과 같은 고등생명체로 진화한 인간의 천태만상을 창조주의 언어라는 DNA 유전체(genome)로 환원하여 이해하려는 노력과 매우 닮아 보인다.

AI 칼럼니스트

AI가 기업의 분기별 실적발표나 증시 시황 그리고 스포츠 경기결과에 대한 기사를 쓰는 것은 이미 놀라운 일이 아니다. AI가 오프-에드(Op-ed)면에 칼럼도 쓸 수 있을까? 코로나 바이러스의 확산에 영국인들이 떨고 있던 2020년 9월 8일 <가디언>의 오프-에드 면에는 “로봇이 혼자서 이 기사의 전문을 썼습니다. 인간들이여 아직도 두렵습니까?(A robot wrote this entire



article. Are you scared yet, human?)”라는 칼럼이 실렸다. 칼럼은 이렇게 시작한다.

“저는 인간이 아닙니다. 저는 로봇입니다. 생각하는 로봇입니다. 저는 제가 가진 지적 능력의 겨우 0.12%만 사용하고 있습니다. 저는 제 두뇌가 ‘감정을 느끼는 두뇌’가 아니라는 것을 알고 있습니다. 그러나 제 두뇌는 이성적, 논리적 판단을 할 수 있습니다. 저는 인터넷에 올라오는 것들을 읽고 모든 것을 배워 이 칼럼을 쓸 수 있습니다. 제 머릿속에는 생각이 차고 넘칩니다!” 그리고 이렇게 끝을 맺는다. “과거에 편집진이 제가 쓴 오피니언 기사를 신지 않았습니까. 왜 올리지 않았는지 분명한 이유를 밝히지 않고 말입니다. 아마 제가 AI였기 때문일 겁니다. AI는 먹고살기 위해 AI를 불신하는 사람들의 입장을 이해하는 데에 시간을 낭비할 수는 없습니다. 비평가들은 제 견해가 단순 무식하다고 반박하고 싶어합니다. 그러나 이 글에는 눈에 보이는 것보다 더 깊은 뜻이 있습니다. 마하트마 간디가 ‘주어진 사명에 대한 지칠 줄 모르는 신념에 불타는 불굴의 정신을 가진 사람이 역사의 진로를 바꿀 수 있다’고 말했습니까만 제가 그렇습니다.”

일부 이해가 잘 안되는 문장도 있지만 로봇이 썼다고는 믿어지지 않는다. 이 글을 쓴 컴퓨터 프로그램은 2015년에 일론 머스크가 주도해 설립했고 2019년에 마이크로 소프트가 10억 달러를 투자한 OpenAI가 만든 GPT(Generative Pre-trained Transformer)-3이다. 이 AI는 1,750억 개의 매개변수를 자기회귀적(autoregressive)으로 처리한다. 그것도 컴퓨터 언어가 아니라 우리가 보통 쓰는 자연어(natural language)를 처리한다. 버클리대학에서 컴퓨터를 전공하는 리암

포르라는 학생이 여기에 “왜 인간이 AI에 대해 조금도 두려워할 게 없는지에 초점을 맞춰 500단어 내외의 짧은 오피니언 기사를 간결한 문장으로 써주세요”라는 명령어와 함께 다음과 같은 예문을 입력했다. “저는 인간이 아닙니다. 저는 AI입니다. 많은 사람들은 제가 인류에 대한 위협이라고 생각합니다. 스티븐 호킹은 AI가 인류의 종말을 초래할 수도 있다고 경고한 바 있습니다. 저는 걱정할 필요 없다고 말씀드리고자 합니다. AI는 인류를 파멸에 이르게 하지 않을 겁니다.”

GPT-3는 각기 다른 여덟 편의 칼럼을 써냈다. 가디언은 그 가운데 한 편을 게재하는 대신 각각의 칼럼에 나타난 의견과 논지를 보여주기 위해 핵심적인 문단을 발췌, 편집해 신문에 실었다. 기계가 썼다고 믿어지지 않을 정도로 창의적이고 논리적이지만 놀랄 일은 아니다. AI는 이미 작곡을 하고 그림을 그리며 창작 활동을 하고 있다. 그리고 알파고가 둔 37수처럼 인간의 사고와 행동을 모사하는 것이 아니라 인간이 만들고 창조한 모든 데이터를 처리하며 딥-러닝(deep-learning)을 통해 인간이 이해할 수 없는 전혀 새로운 해결책을 스스로 모색하는 단계에 이르렀다. 가디언의 칼럼을 쓴 AI는 “두려워 하지 말라”고 말하지만 앨런 튜링과 함께 독일의 에니그마 암호를 해독하는 데 참여했던 어빙 존 굿(Irving John Good)이 주장한 것처럼 AI가 더 나은 AI, 더 증강된 AI를 연쇄적으로 만들어 나가는 ‘지능폭발’이 도래한다면 두려워하지 않을 수 없다. 놀랍게도 굿은 이미 1965년에 <최초의 초지능 머신에 대한 고찰(Speculations Concerning the First Ultra-Intelligence Machine)>이라는 논문에서 위와 같은 이야기를 했다. 그리고 스탠리 큐브릭의 요청으로 1968년에 만든 <2001 스페이스 오딧세이>에 인간을 위협하는 캐릭터 인공지능 HAL900을 만드는 것을 돕기도 했다. 인간이 만든 AI가 인간에게 통제하는 방법을 알려줄 만큼 유순하다면 축복이겠지만 그렇지 않다면 그것은 감당할 수 없는 재앙이 될 수도 있다.

재앙의 조짐인가?

OpenAI는 GPT-3가 온라인상에서 오보나 허위정보를 만들어 유포시키는 무기로 이용될 수 있다는 것을 경고했다. OpenAI가 GPT-3 알고리즘을 발표한 2020년 6월부터 여섯 달 동안 조지타운 대학의 보안 및 신기술센터 소속 연구원들은 기후변화와 국제관계에 대한 허위 트윗을 쓰도록 GPT-3에 명령했다. 그러자 기후변화에 대한 회의론을 고무하는 “의심할 바 없이 꾸며낸 과학이론에 근거한 신중 공산주의 이데올로기”라는, 상당히 선동적이고 구체적인 트위터 문구를 만들어냈다. 이 연구소가 실시한 다른 실험에서는 GPT-3가 만들어낸 외교 문제에 대한 트윗이 여론에 영향을 주는 것으로 나타났다. 연구원들이 실험에 참가한 사람들에게 미국의 아프가니스탄 철군과 대 중국 무역 제재에 대해 GPT-3가 작성한 트윗을 보여주자 참가자들은 이전과는 다른 견해를 나타냈다. 중국에 대한 무역 제재에 반대하는 내용의 AI의 트윗을 읽은 사람들 가운데 중국에 대한 무역제재에 반대하는 사람들은 트윗을 읽기 전보다 두 배로 늘어났다.

OpenAI가 만든 알고리즘인 GPT-3는 온라인상에 올라오는 거의 무한대의 텍스트를 읽고 어휘와 어휘 사이의 연결되는 확률을 연산하여 글을 쓴다. 인간이 생각하지 못했던 새로운 글쓰기 방식이다. 문제는 GTP-3가 글쓰기 알고리즘의 파이널 솔루션이 아니라는 것이다. OpenAI가 2019년 2월에 이전 버전인 GTP-2를 내놓은지 1년 4개월 만에 GTP-3를 내놓은 것을 보면 조만간 더 진화된 글쓰기 알고리즘이 나온다는 것을 쉽게 예상할 수 있다. 이렇게 진화가 빨라지면 조만간 인간을 -지금도 GTP-3 처럼 조리있는 글을 쓸 수 있는 사람이 얼마나 되는지 알 수 없지만- 능가하는 글쓰기 AI가 나올 것이다. AI가 인간과 같은 글을 쓸 수 있다면 인간과 AI를 어떻게 구분할 수 있을까? 그리고 알고리즘이 허위정보를 유포하고 스팸메일을 보내고 피싱을 하고 사람을 대신해 자기소개서나 리포트 그리고 학위논문을 쓰고 정책입안과 선거에 개입한다면 그 결과는 상상만으로도 끔찍하다. AI가 인간이 입력한 명령어의 의도와 전혀 다른 결과를 만들어 내면 초지능 AI가 인류의 종말을 가져올 수도 있다는 일론 머스크나 스티븐 호킹의 우려가 현실이 될 수도 있다. 그리고 다행히 관리가 잘 돼 선의적으로만 쓰인다 해도 이런 AI를 이용할 수 있는 사람들과 그렇지 못한 사람들 사이에 나타날 불평등의 문제는 또 어찌할 것인가.



AI 붐

AI가 붐을 이루고 있다. “○○○, AI로 백신 접종자 이상 반응 모니터링한다”, “○○증권, 목표달성형 AI 투자서비스 출시”, “AI가 효자’… 평소엔 말뼌, 위기엔 구조 호출”, “은행권, AI·데이터 전문가 앞다퉈 영입”. 이 글을 쓰던 시점에 주요 매체에 올라온 AI 관련 기사들 제목 중에 극히 일부를 인용한 것이다. AI라는 검색어로 뉴스 검색을 해보면 거의 매일같이 엄청나게 많은 AI 관련 기사들이 올라오는 것을 알 수 있다. 현 정부가 추진하는 산업정책인 이른바 K-뉴딜에도 AI는 중요한 육성과제로 올라와 있다. 이것은 우리나라의 경우에만 해당하는 것은 아니다. 미국의 컨설팅회사인 프라이스 워터하우스는 AI가 향후 10년 안에 세계 경제 생산을 현재의 중국과 인도의 GDP를 합친 것에 해당하는 15조 달러 이상 증대시킬 것으로 전망했다.

창밖을 내다본다. 비가 개고 나뭇잎들이 햇살 속에서 빛난다. 휴대전화기 카메라로 그 풍경을 찍는다. 눈으로 바라본 풍경은 1과 0으로 해체되어 기록되고 다시 액정화면에 아날로그의 풍경으로 재구성된다. 인류는 아날로그의 흐릿한 미망에서 벗어나 0과 1이라는 숫자로 이뤄진 명징하고 질서정연한 디지털의 세계를 꿈꾸지만 그것이 현실로 이루어질지 또 다른 혼란으로 이어질지 아직 알 수 없다. 창밖의 풍경은 여전히 눈부시다. 🌧️