

홍성욱 한국천문연구원 선임연구원

## 요 약

인공지능, 혹은 기계학습은 현대 과학기술에서 가장 관심을 받는 분야 중 하나이다. 기술 분야뿐 아니라, 자연의 기본 원리를 이해하고자 하는 자연과학 분야에서도 인공지능이 널리 적용되고 있다. 하지만 인공지능을 이용한 연구를 통해 인류가 정말로 자연을 제대로 이해할 수 있는가에 대한 우려도 크다. 이러한 우려의 예로서는 (1) 인공지능이 아직 일어나지 않은 현상에 대해서 예측을 줄 수 있는가, (2) 인공지능이 얻은 결과의 정확성을 통계적으로 이해할 수 있는가, (3) 인공지능을 통해 자연법칙 발견 혹은 현상에 대한 물리적인 이해를 얻을 수 있는가 등이 있다. 본 발표에서는 위와 같이 인공지능이 자연과학 분야에 적용되는 사례와 우려에 대해 간단히 알아보고, 이러한 논의가 기독교인이 자연과학을 바라보는 태도에 어떠한 영향을 줄 수 있는지도 생각해 본다.

주제어: 인공지능, 기계학습, 자연과학

## I. 서 문

오늘날의 사회, 혹은 사회가 나아가야 할 방향을 가리키는 키워드 중 하나가 ‘4차 산업혁명’이다. 4차 산업혁명은 독일에서 2010년부터 연구된 개념으로, 증기 엔진의 개발과 함께 산업화가 이루어진 1차 산업혁명(1760~1820년), 전기, 전신, 열차 등을 통한 산업 확장이 이루어진 2차 산업혁명(1870~1914년), 컴퓨터의 개발과 함께 산업의 디지털화가 이루어진 3차 산업혁명(1980년대~) 다음에 올 사회상을 뜻한다. (김은, 2017; Berlanstein, 2003; Engelman, 2015; Janssen) 4차 산업혁명이 가지는 주요 특징으로는 상호연결, 정보의 투명성, 인간의 문제해결과 의사결정 과정에서 기술의 적극적 개입, 의사결정의 탈중양화 등이 있으며, 이

과정에서 가상물리시스템, (산업용) 사물인터넷, 클라우드 컴퓨팅, 인공지능 등의 기술이 새로 주목받게 된다고 한다. (Hermann, Pentek & Otto, 2016)

4차 산업혁명과 관련된 다양한 기술 중, 대중의 관심과 우려를 동시에 불러일으키는 개념 중 하나가 바로 인공지능(artificial intelligence)이다. 인공지능은 인간이나 동물이 갖고 있는 자연지능이 아닌, 기계가 가지는 지능을 의미한다. (Poole, Mackworth & Goebel, 1998; Russel & Norvig, 2003) 인간이 만들어낸 피조물이 지성을 갖게 되는 것에 대한 상상은 훨씬 예전부터 있어왔는데, 중세 연금술사들이 상상한 호문쿨루스나 소설 ‘프랑켄슈타인’에 나오는 괴물이 그 대표적인 예라고 할 수 있다. (Paracelsus, 1537; Shelly, 1818) 이후 컴퓨터가 개발되면서 인공지능의 가능성이 열리기 시작하였으며, 하드웨어와 소프트웨어의 눈부신 발달로 인해 기존에는 인간만이 수행할 수 있다고 생각했던 문제의 상당수를 이제는 기계가 더 우월하게 수행하게 되었다. 특히 1996년 IBM의 딥블루(Deep Blue)가 체스 세계 챔피언을 이긴 사건이나, 2016년 구글의 알파고(AlphaGo)가 이세돌 9단을 바둑으로 이긴 사건 등은, 공상과학에서만 다루던 강력한 인공지능이 현실에 임박했다는 인상을 대중에게 심어주었다.

특히 2010년대 이후 인공지능의 흐름에서 중요한 역할을 차지하고 있는 것은 기계학습(machine learning), 특히 인공신경망(artificial neural network)이다. 기계학습은 주어진 자료를 처리하는 방법을 인간이 기계에게 전달하는 것이 아니라, 좀 더 느슨한 컴퓨터 알고리즘을 통해 기계가 직접 찾는 것을 의미한다. (Alpaydin, 2020) 기계학습이 할 수 있는 일로는 자료를 주어진 카테고리에 맞게 분류하는 일, 자료로부터 주어진 특정 변량을 예측하는 일(회귀; regression), 자료를 임의의 카테고리로 분류하는 일(군집화; clustering), 주어진 환경과 상호작용하며 최적의 전략을 찾는 일(강화학습; reinforcement learning) 등이 있다. 인공신경망이란 기계학습의 한 종류로서, 인간을 포함한 동물의 뇌에 있는 신경망과 같은 논리 구조를 겹겹이 쌓아서 기계학습을 수행하는 것이다.

오늘날 기계학습을 이용한 인공지능<sup>1)</sup>은 산업 전반에 걸쳐서 응용 방안이 연구되고 있으며, 한국에서도 유수의 기업이 자신의 사업에 인공지능을 적극적으로 도입하고 있다. 산업뿐만 아니라, 자연의 기본 원리를 규명하는 것을 목적으로 삼는 자연과학에서도 인공지능이 다양하게 시도되고 있다. 특히 검색해야 할 자료의 수가 너무 많거나, 전통적인 계산 방식으로는 너무 많은 계산량을 필요로 하는 대기, 생화학 등의 분야에서 인공지능이 활발하게 사용되고 있다. (Wei, 2019) 필자가 연구하는 천문학은 다른 자연과학 분야에 비해서는 상대적으로

---

1) 이 이후로 ‘인공지능’이란 용어는 특별한 언급이 없는 이상 인공신경망과 동의어로 사용한다.

인공지능의 도입이 늦은 편이지만, 역시 2010년대 중반 이후로는 다양한 세부 분야에 대해 인공지능을 적용하려는 움직임이 있다.

그러나 산업 분야(그리고 일부 자연과학 분야)에서 인공지능이 거둔 놀라운 성과에도 불구하고, 상당수의 자연과학 학계 종사자들은 인공지능이 항상 자연과학의 실질적인 발전, 즉 자연의 이해에 도움이 될 것인지에 대해 의문을 품고 있다. 물론 역사적으로 컴퓨터 등 기계를 적극적으로 활용한 자연과학 연구는 항상 당대에 그 진정성에 대해 의심을 받아왔었다. 하지만 인공지능이 유독 더 심한 의심을 받게 되는 것은 인공지능의 동작 방식이 인간에게는 일종의 ‘블랙박스’처럼 여겨지기 때문일 것이다.

본 논문의 II장에서는 인공지능, 특히 심층인공신경망(deep neural network)의 기본 동작 원리를 살펴보고, 인공지능이 자연과학, 특히 천문학에서 어떻게 적용되고 있는지를 살펴본다. III장에서는 자연과학에 인공지능을 적용하는 것에 대한 다양한 우려를 소개한다. 마지막으로 IV장에서는 앞장에서 살펴본 논의가 기독교인이 자연과학을 바라보는 태도에 어떠한 영향을 줄 수 있는지를 간략히 살펴본다.

## II. 인공신경망의 원리와 자연과학에의 적용

### 1. 인공신경망의 역사

인공신경망의 기원은 1943년 신경의학자인 맥컬록(Warren McCulloch)과 논리학자인 피츠(Walter Pitts)가 생물의 신경망(뉴런)이 동작하는 원리를 포현한 논리구조 모형을 수립하고, 이를 1949년 헵(Donald Hebb)이 수정한데서 시작한다. (McCulloch & Pitts, 1943; Hebb, 1949; 그림 1 왼쪽) 이들이 제시한 신경망의 작동 원리에서 주목할 점은, 신경망의 출력값은 입력값의 변화에 선형적으로 반응하지 않으며, 오히려 입력값이 일정 크기 이상이 되기 전에는 출력값에 제대로 반영되지 않는다는 점이다. 1957년 심리학자인 로젠블랫(Frank Rosenblatt)은 앞선 신경망의 논리구조 모형을 반영한 최초의 인공신경망인 퍼셉트론(Perceptron)을 개발하여, 뇌가 시각 신호를 처리하는 것을 흉내내려 하였다. (New York Times, 1958; 그림 1 오른쪽)

1966년 수학자인 이바넨코(Alexey Ivakhnenko)는 입력-출력 사이에 한 수준 이상의 인공 신경망을 쌓아, 실질적으로 최초의 심층기계학습을 구현하였다. (Ivakhnenko, 1871) 그리고 1979년 컴퓨터과학자인 후쿠시마(Kunihiko Fukushima)는 오늘날 이미지를 이용한 인공지능

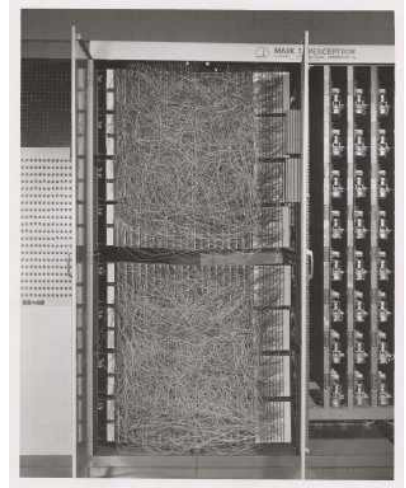
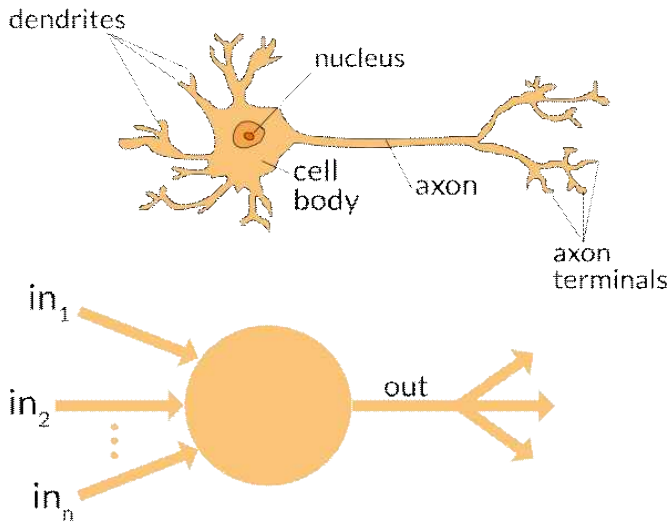


그림 1. (왼쪽) 동물의 신경망(뉴런)과 그 작동 원리를 표현한 논리구조 모형.

(<https://appliedgo.net/perceptron/>) (오른쪽) 최초의 인공신경망인 퍼셉트론. (Cornell Library)

에서 널리 사용되는 합성곱(convolution)을 사용한 최초의 인공신경망을 구현하였다. (Fukushima, 1979) 1982년 홉필드(John Hopfield)는 여러 겹의 신경망에서 전후 단계의 신경망에서 나오는 출력값을 다른 단계의 입력값에 포함시킴으로써, 초기 단계의 정보가 신경망을 거치면서 손실되는 것을 막는 기법을 개발했다. (Hopfield, 1982) 1991년 슈미트후버(Jurgen Schmidhuber)는 1000겹 이상의 인공신경망이 당시 컴퓨터에서 작동한다는 것을 보여주었으며, 이후 2011년 크리제프스키(Alex Krizhevsky)는 ImageNet에서 주최한 이미지 분류 알고리즘 경연대회에서 합성곱 인공신경망을 적용하여 우승함으로써, 인공신경망이 산업계의 주목을 받는데 큰 기여를 하였다.

## 2. (심층)인공신경망의 주요 특징

그림 2의 왼쪽 그림은 심층인공신경망의 기본 개념을 나타낸다. 세로로 이루어진 줄을 '레이어'(layer)라고 부르며, 각 레이어는 여러 개의 변수를 포함한다. 서로 다른 레이어의 변수를 연결하는 선(뉴런)은 한 단계의 변수가 다른 단계의 변수의 변화에 영향을 주는 비중으로 정의된다. 이 때 심층인공신경망은 주어진 입력 자료에 대해 최대한 출력 자료와 가까운

자료를 내도록 각 뉴런에 부여된 비중을 조절한다.

(심층)인공신경망 이전에도 컴퓨터를 이용하여 주어진 문제에 최적화된 모형을 찾는 일은 많이 이루어져 왔다. 하지만 기존의 최적화 기법에 비해 심층인공신경망에는 다음과 같은 세 가지 중요한 특징이 있다. 첫 번째로, 기존의 최적화 기법은 입력과 출력을 연결하는 하나의 정해진 형태의 함수를 찾으려 하는데 반해, 심층인공신경망은 입력과 출력 사이에 여러 개의 숨겨진 레이어를 사용하는 경향이 있다. 숨겨진 레이어를 사용하여, 입력 자료에 존재하는 특징을 다양한 방법으로 끄집어내는 것을 목표로 한다. 두 번째로, 심층인공신경망에서는 비선형적 함수를 적극 활용한다. 만약 모든 뉴런이 선형대수 함수로만 이루어진다면, 입력과 출력 사이에 아무리 많은 레이어를 적용하더라도 이와 정확히 같은 결과를 주는 입력-출력으로만 이루어진 알고리즘을 얻을 수 있다. 즉, 선형함수만 이용할 경우 숨겨진 레이어를 활용하여 얻는 이득이 없다. 이를 해결하기 위해 추가로 적용하는 비선형적 함수를 ‘활성화 함수’(activation function)라고 부른다. (그림 2 오른쪽) 마지막으로, 심층인공신경망에서는 숨겨진 레이어를 적용하여 얻은 수많은 정보량 중 일부를 버리는 과정을 적용하는 경향이 있다. 이러한 과정을 거치지 않을 경우 심층인공신경망에 들어 있는 변량의 수가 입출력 자료의 수보다 많아질 수 있으며, 주어진 자료에 들어 있는 잡음을 (우리가 찾고자 하는) 전체적인 경향성과 구별하지 못하는 과적합(overfitting) 현상이 일어나게 된다. 이를 방지하고자 일부 뉴런을 사용하지 않거나, 뉴런에 들어 있는 비중을 재규격화하는 등 정보량을 줄이는 일을 추가한다. (Ioffe & Szegedy, 2015)

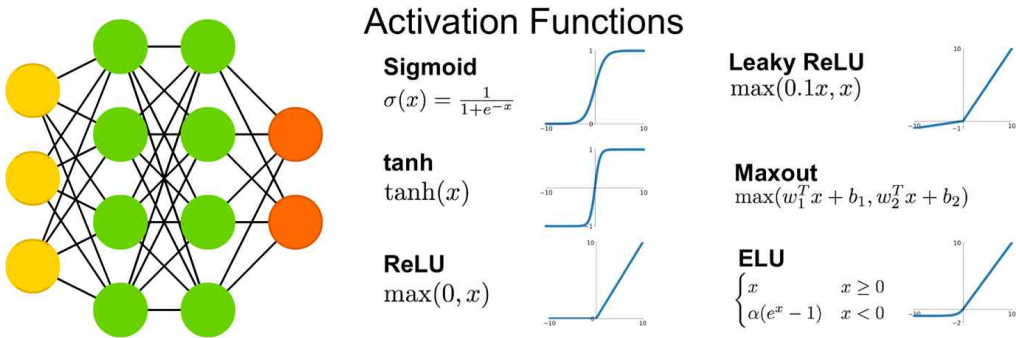


그림 2. (왼쪽) 심층인공신경망의 기본 개념도. 입력(노란색)과 출력(빨간색) 사이에 숨겨진 레이어(초록색)가 있다. (오른쪽) 인공신경망에 비선형성을 부여하는 활성화함수의 예시.

(<https://medium.com/@shrutijadon10104776/survey-on-activation-functions-for-deep-learning-9689331ba092>)

### 3. 자연과학에의 적용 예시

I장에서 이미 언급하였듯이, 인공지능은 처리해야 할 자료의 양이 많거나, 물리법칙을 이용한 전통적인 계산 방식으로는 계산량이 너무 많은 문제의 경우에 강점을 보인다는 것이 알려져 있다. 자연과학의 여러 문제 중 이와 같은 특성을 가진 문제, 즉 기후 예측이나 3차원 단백질 구조 예측에서 인공지능이 적극적으로 도입되었다. 특히 3차원 단백질 구조 예측의 경우 이세돌 9단을 바둑에서 꺾은 알파고(AlphaGo)를 만든 구글 딥마인드(DeepMind)가 알파폴드(AlphaFold)라는 인공지능을 제시했고, 이것이 CASP라는 단백질 구조 예측 경연대회에서 기존 전문가그룹을 꺾고 우승하였다. (Senior *et al.*, 2020)

필자가 연구하고 있는 천문학의 경우, 1차 관측 자료는 곧 하늘에서 오는 천체의 빛이다. 기술이 발전하면서 더 좋은 망원경과 장비를 이용하여 효율적으로 천체의 빛을 모으고 분석하려 하지만, 기본적으로 천체에서 오는 신호에는 천체와 관찰자 사이에 존재하는 다양한 환경에 의한 효과가 존재한다. 또한, 실험실에서 환경을 통제하면서 연구자가 관심있어 하는 효과만을 살펴볼 수 있는 다른 자연과학과는 달리, 천문학에서는 연구자가 천체와 주변 환경을 통제할 수 있는 자유도가 매우 적다. 즉, 천문학에서 다루는 현상의 기본 원리는 매우 단순하더라도, 앞서 설명한 다양한 효과 때문에 실제로 일어난 현상을 계산으로 모사하는 일은 대체로 어렵다.

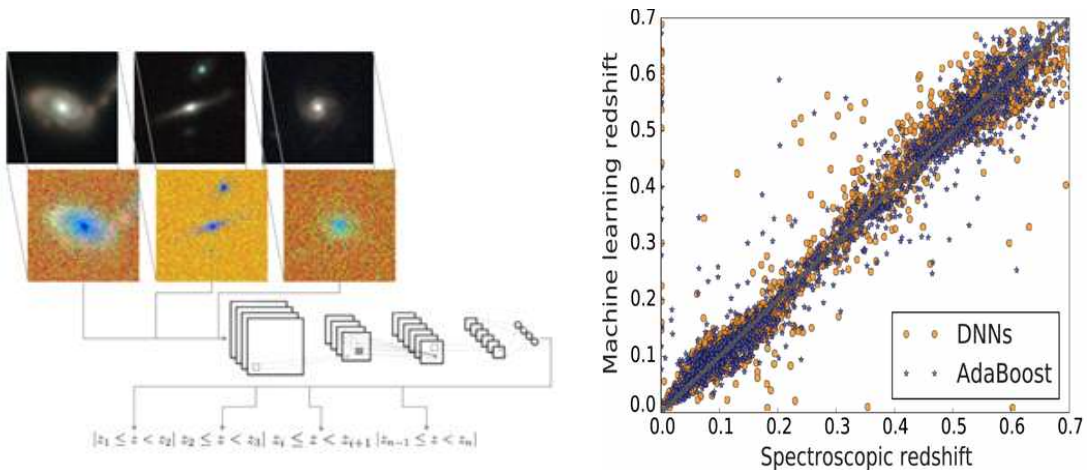


그림 3. 합성곱 심층인공신경망을 이용한 측광 적색이동 예측 방법(왼쪽)과 그 결과(오른쪽). (Hoyle, 2016)

2010년대 중반부터 천문학에 인공지능, 특히 심층인공신경망이 도입되기 시작하였는데, 그 대표적인 적용사례는 측광 적색이동(photometric redshift) 예측과 은하의 형태(morphology) 분류이다. 우선 적색이동은 도플러 효과(Doppler effect)에 의해 멀어지는 물체에서 나오는 파장이 우리에게는 원래보다 길어지게 보이는 정도를 나타내는 값이다. 현대 우주론에 의하면 우주 공간은 팽창하고 있으며, 멀리 있는 천체가 더 빠른 속도로 우리에게서 멀어지게 된다. 따라서 적색이동을 측정하면 천체까지의 거리를 간접적으로 알 수 있으며, 이 정보는 우리 우주의 전체 구조를 이해하는데 큰 도움이 된다. 적색이동을 측정하기 위해서는 원칙적으로 분광 관측을 통해 천체의 스펙트럼을 얻어야 하나, 멀리 있거나 어두운 천체의 경우 스펙트럼 관측이 매우 어렵다. 반면 필터를 적용한 뒤 카메라로 촬영하는 측광 관측의 경우 어두운 천체도 상대적으로 쉽게 관측할 수 있지만, 스펙트럼의 선을 직접 보는 것이 아니기 때문에 적색이동의 정밀도가 매우 낮아진다. 하지만 최근에는 인공지능을 이용하여 측광 적색이동의 정확도와 정밀도를 향상시키고자 하는 연구가 많이 이루어지고 있다. (Carrasco Kind & Brunner, 2014; Hoyle, 2016; 그림 3).

은하의 형태는 크게 나누면 타원형 은하(elliptical galaxy)와 나선형 은하(spiral galaxy)로 나눌 수 있으며, 막대의 유무나 나선의 개수, 둥그러운 정도 등에 따라 더 다양한 세부 분류로 나눌 수 있다. 은하의 형태는 은하가 만들어진 뒤 어떠한 과정을 거쳐 왔는지를 나타내는 가장 중요한 단서이며, 은하의 다양한 물리량을 계산할 때 은하의 형태가 중요한 역할을 한다. 전통적으로는 숙련된 천문학자가 은하를 촬영한 사진을 하나씩 살펴보며 은하의 형태를 분류했다. 하지만 앞으로는 단시간 내에 수백만 개의 천체를 촬영할 수 있는 천문관측 프로젝트가 많이 이루어질 예정이며, 이를 대비하기 위해서는 은하 형태 분류를 자동화할 수 있는 방안이 마련되어야 한다. 기존에는 앞서 소개한 은하의 형태 분류에 필요한 요소를 함수의 형태로 변환하기 어려웠지만, 합성곱 심층인공신경망이 이미지 분류에 탁월하다는 것이 알려지면서 인공지능이 은하 형태 분류에 본격적으로 활용되고 있다. (Dieleman, Willet & Dambre, 2015)

이외에도 인공지능은 은하의 분포를 자동으로 통계분석하여 우주를 기술하는 기본 상수의 값을 예측하거나, (Pan *et al.*, 2020) 고해상도로 수행된 유체역학 시뮬레이션에서 얻은 결과를 저해상도 다체 시뮬레이션에 덮어씌워 은하의 물리적 특성을 잘 반영하면서도 큰 부피를 지닌 시뮬레이션 자료를 생성하거나, (Jo & Kim, 2019) 기존에 관측한 천체의 이미지로부터 아직 사용하지 않은 파장의 천체 이미지를 예측하는 연구 등도 이루어지고 있다. (Kim *et al.*, 2019)

### III. 인공지능은 자연의 이해에 도움이 되는가?

앞에서 살펴 보았듯, 오늘날에는 자연의 기본 원리를 이해하고자 하는 자연과학의 다양한 분야에 인공지능이 적용되고 있다. 이렇게 인공지능을 자연과학에 적극적으로 도입하는 연구자들은, 인공지능은 기존의 이론이나 계산법으로는 풀기 어렵거나 불가능했던 문제를 손쉽게 풀 수 있기 때문에 자연과학의 발전에 도움이 된다는 입장을 갖고 있다. 더 나아가서, 인공지능은 인간의 지성으로는 인지하기 어려운 새로운 패턴을 발견할 수 있기 때문에, 새로운 과학적 발견을 이끌어 낼 수 있으리라는 기대를 갖고 있다.

하지만 인공지능에 대한 기대에 못지 않게, 인공지능을 자연과학에 도입하는 것에 대해 우려를 표하는 과학자들도 매우 많다. 이러한 우려 중에는 인공지능의 도입으로 인해 컴퓨터공학자나 데이터과학자 등 전통적인 자연과학 전문가의 범위에서 벗어난 사람들이 무분별하게 학계에 유입되어 영향을 끼칠 것에 대한 걱정과 같은 사회적인 측면도 많이 존재한다. 하지만 그보다 훨씬 근본적인, 정말 인공지능이 인류가 자연의 기본 원리를 이해하는데 근본적인 도움을 줄 수 있을 것인가에 대한 우려도 상당히 많다. 본 장에서는 이러한 우려를 몇 가지로 정리해 보고, 각각에 대해서 논의해 보도록 한다.

## 1. 인공지능은 아직 일어나지 않은 현상에 대해 예측할 수 있는가?

현대 자연과학이 철학, 종교 등 자연을 설명하는 다른 종류의 학문에 비해 보편적으로 더 설득력을 갖는다고 여겨지는 가장 중요한 이유는, 자연과학은 이미 일어난 일 뿐 아니라 앞으로 일어날 일에 대해서도 정량적인 예측을 해줄 수 있으며, 예측과 실제 현상을 관측 비교함으로써 특정 자연과학 모형의 옳고 그름을 (연구자의 경향과는 어느 정도 독립적으로) 정량적으로 점검할 수 있기 때문일 것이다. 따라서 인공지능이 자연과학의 발전에 근본적인 도움이 되려면, 특정 현상에 대해 수립한 인공지능 모형이 아직 일어나지 않은 현상 역시 잘 예측할 수 있어야 할 것이다.

자연과학에서 인공지능을 도입할 때는 주로 분류나 회귀와 같은 지도학습(supervised learning)을 적용하는데, 보통 지도학습을 수행할 때는 인공지능 모형이 일반적인 현상에 대해 충분한 예측력을 줄 수 있도록 여러 조치를 취하게 된다. 그중 가장 대표적인 것은 인공지능 모형을 수립하기 위해 연구자가 사용하는 자료를 여러 종류로 분할하는 것인데, 보통 훈련 표본(train sample), 유효판별 표본(validation sample), 검정 표본(test sample)로 분할한다. 훈련 표본은 인공지능 모형 훈련에 직접 사용되어, 레이어간 연결선의 비중을 수정하는데 사용된다. 유효판별 표본은 훈련 과정의 매 단계마다 사용되며, 주어진 단계까지 훈련 표본을



통해 조절된 인공지능 모형이 훈련 표본이 아닌 자료에 대해서도 예측력을 줄 수 있는지를 점검하고, 연구자가 훈련 과정을 언제까지 지속할지 판단하는데도 사용된다. 마지막으로 검정 표본은 훈련 과정이 완전히 끝난 뒤에 사용되며, 최종적으로 얻어낸 인공지능 모형이 훈련/유효 판별 표본이 아닌 자료에 대해서도 어느 정도의 예측력을 줄 수 있는지를 연구하는데 사용된다. 사실 자료를 훈련/유효판별/검정 표본으로 분할하는 것은 인공지능 뿐 아니라 통계분석에서 널리 쓰이는 방법이다. (James *et al.*, 2013) 따라서 기존에 통계적 방법론을 사용하던 자연과학 문제에 대해 앞의 과정과 같이 인공지능을 적용한다면, 기존 통계분석에 비해 인공지능에만 예측력을 의심하는 것은 부당하다고 할 수 있다.

사실 일어나지 않은 현상에 대해 충분한 예측력을 줄 수 있느냐는 ‘인공지능을 적용한 자연과학’만의 문제가 아니라 ‘자연과학’ 그 자체에 해당하는 문제이다. 많은 자연과학자들은 어떠한 자연현상이든지 주어진 현상을 정량적으로 설명하는 근본 이론이나 시나리오가 실제로 존재한다고 믿으며, 또한 우리가 자연을 충분히 관측하면 이 근본 이론과 시나리오에 충분히 가까운 모형을 실제로 수립할 수 있다고 믿는다. (Heilbron, 2003; Chen, 2009) 아직까지는 자연과학자들의 믿음과 같이 현대 자연과학이 아직 일어나지 않은 사건에 대해서도 충분한 예측력을 제공해 준 것으로 보이고, 그 덕분에 자연과학을 통한 문명 발전이 일어날 수 있었다. 하지만 우리 우주에 존재하는 모든 자연현상에 대해 반드시 근본 이론이나 시나리오가 존재해야 할 논리적인 근거는 없다. 마찬가지로 인간이 수행할 수 있는 유한한 관측에서 도출해 낸 내삽(interpolation)과 외삽(extrapolation)이 항상 잘 맞으리라는 논리적인 근거도 없다. (Hume, 1739)

## 2. 인공지능이 얻은 결과의 정확성을 통계적으로 이해할 수 있는가?

기존에 널리 사용되던 통계모형의 경우, 어떤 현상에 대해 예측값을 줄 뿐만 아니라 기존의 학습을 통해 실제와 예측 사이에 있는 오차를 어느 정도로 예상할 수 있는지도 알 수 있다. 또한 주어진 문제에 대해서 통계모형을 사용할 때 결과가 항상 일정한 값으로 수렴할지의 여부도 미리 알 수 있다. 이러한 성질은 가우스 과정(Gaussian process)을 포함한 많은 기계학습 모형에도 잘 적용되지만, 유독 심층인공신경망에 대해서는 위와 같은 통계적 성질이 제대로 규명되지 않았다고 여겨져 왔다. 이러한 상황에서 심층인공신경망이 주는 예측값의 오차나 수렴성 등을 알기 위해서는, 대부분 각 문제에 대해 유한한 검정 표본에 직접 실험하는 데 크게 의존해 왔다.

하지만 최근에는 심층인공신경망을 통해서도 통계적 성질을 얻기 위한 노력이 계속 이루어지고 있다. 예를 들어 Allen-Zhu, Li & Song (2019)에서는 심층인공신경망의 뉴런의 개수가 훈련 표본의 수에 비해 충분히 큰 경우 인공지능이 주는 결과는 반드시 수렴한다는 것을 보였다.<sup>2)</sup> 또 심층인공신경망이 항상 주어진 문제의 광역 최적화(global optimization)에 근접할 것을 담보하는 최적화 기법이 계속해서 제안되고 있다. (Wang et al., 2019; Reddi, Kale & Kumar, 2019; Yi, Ahn & Ji, 2020) 또한 문제의 예측값 뿐 아니라 오차에 대한 자세한 정보를 제공해 주는 공분산 행렬(covariance matrix)의 고유값(eigenvalue)을 출력값으로 주는 심층인공신경망도 연구되고 있다. (Liu et al., 2018)

### 3. 인공지능을 통해 자연법칙을 발견하거나 물리적인 이해를 줄 수 있는가?

인공지능을 통해 기존의 자연법칙이나 전통적인 계산 방법으로 풀기 불가능하거나 어려운 문제를 손쉽게 풀 수 있다고 하더라도, 그것이 (특히 인간에게) 새로운 자연법칙의 발견이나 물리적인 이해를 줄 수 있을지에 대해서는 많은 자연과학자들이 긍정적으로 바라보지 않고 있다. 이에 대한 가장 큰 원인은 자연과학이 단지 주어진 자연 현상을 기계적으로 재현하는 것에서 그치는 것이 아니라, 그 자연 현상이 주어진 방식으로 일어나게 되는 ‘근본적인 원리’를 ‘인간의 지성으로’ 이해하는 것이기 때문이다. 전통적으로 자연과학자들은 주어진 자연 현상을 설명하는 함수를 해석함수와 같이 그 모양이 잘 알려진 함수를 이용해 왔다. 특히 함수의 모양은 단지 좋은 예측값을 주는데만 사용될 뿐 아니라, 그 자체로 자연과학자에게 해당 자연 현상이 동작하는 원리에 대한 물리적 통찰력을 제공해 줄 수 있다.

반면, 인공지능은 간단한 함수의 형태로 나타내는 것이 불가능하며, 따라서 전통적인 자연과학자의 입장에서는 거의 ‘블랙박스’와 다름없이 동작하게 된다. 물론 학습이 끝난 인공지능 모형이 어떤 식으로 동작하는지를 간접적으로 알아보는 것은 가능하며, 주어진 인공지능 모형이 동작하는 방식을 시각적으로 표현하기 위한 다양한 방법이 개발되어 있다. (Springenberg et al., 2014; Yoshinki et al., 2015; Zhou et al., 2016; 그림 4) 하지만 이러한 시각화는 주로 연구자가 인공지능 모형의 동작 원리를 정성적으로 짐작하는데 도움이 될 뿐이며, 전통적인 자연과학자가 추구하는 것과 같은 현상에 대한 완전한 통찰을 이끌어주기는 어렵다.

---

2) 주의할 점은, 이 논문에서 결과가 수렴하는 현상은 최적화에 들어가는 변수의 수가 사용하는 자료의 수보다 많아 과적합(overfitting)이 일어나기 때문이다. II.2장에서 살펴보았듯이 과적합 현상이 일어난 인공지능 모형은 일반적인 현상을 예측하는 능력이 오히려 떨어지게 되며, 따라서 이러한 모형은 실제 상황에서 활용될 수 없다.

더 근본적인 문제는, 위와 같은 행위는 결국 자연 현상의 근본적인 원리를 이해하기 위해 인간이 준비한 인공지능 모형의 근본적인 동작 원리를 인간이 완전히 이해하지 못한다는 점을 시사하고 있다. 우리가 탐구하고자 하는 자연 현상에는 근본적인 원리가 숨어 있을 수 있고, 어쩌면 우리가 마련한 인공지능이 아주 복잡한 심층신경망 구조를 이용하여 그 근본적인 원리를 어떠한 식으로든 파악했을 지도 모른다. 하지만 우리가 정작 인공지능 모형에 기록되어 있는 수많은 뉴런 구조와 비중값으로부터 인공지능이 동작하는 원리에 대한 정량적 통찰을 얻지 못한다면, 우리는 애초에 원하던 자연 현상의 근본적인 원리의 통찰을 얻지 못하는 것이 아닌가? 좀 더 극단적으로 말하자면, 우리가 자연과학을 이해하는데 인공지능을 더 적극적으로 사용하면 할수록, 기계가 인간보다 자연에 대한 더 깊은 이해를 갖게 하는 실수를 범하는 것이 아닌가? (Shanahan, 2015)

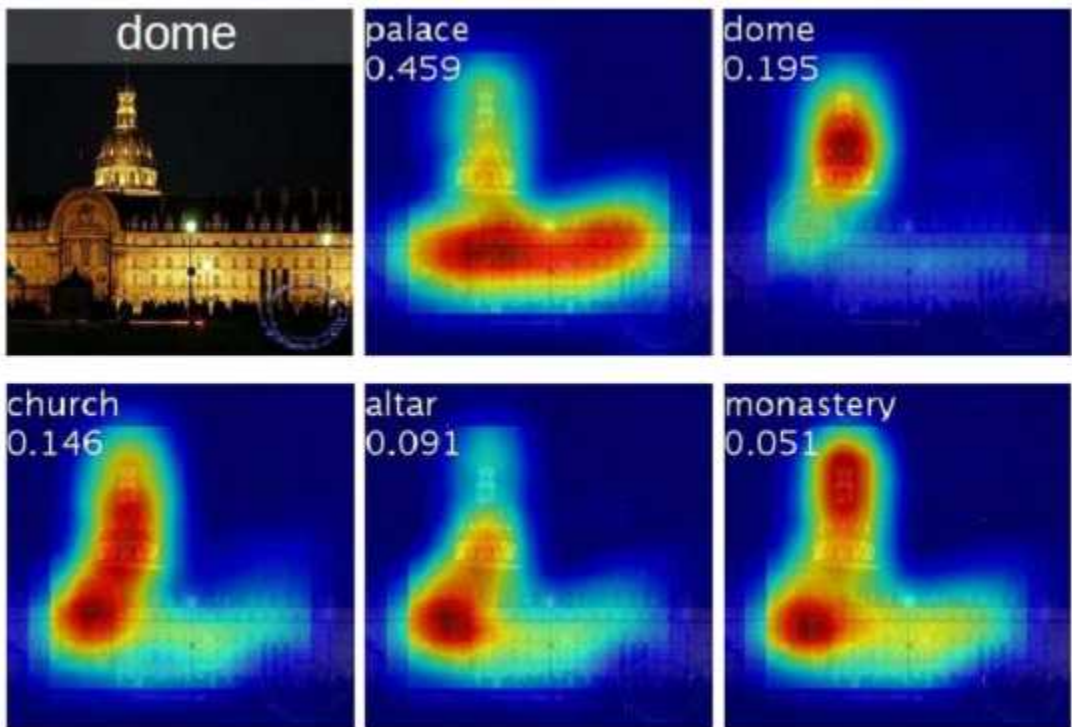


그림 4. 이미지를 설명하는 표제어를 출력하는 특정 인공지능 모형이 각 표제어를 예측할 때 주로 참고한 이미지 영역의 중요도를 나타낸 그림.

특정 문제를 풀기 위해 도입한 인공지능이 어떻게 동작하는지를 이해하기 위해서는 인공지능 모형 자체에 대한 실험 관찰이 필요하다. (Zhou et al., 2016)

#### IV. 기독교인으로서 생각해 볼 점

앞선 자연과학과 인공지능에 관한 여러 논의를 살펴보면, 필자가 기독교인으로서 느낀 점 두 가지를 아래와 같이 정리해 보았다.

첫 번째로, 오늘날 자연과학을 포함한 사회 전반에서 인공지능이 두각을 드러내는 현상은, 오히려 하나님이 이 세상을 경영하는 방식이 인간의 통찰력으로 완전히 이해하기 어렵다는 것을 드러내 주는 것 같다. 17세기부터 시작된 과학혁명 이후로 우주의 모든 사물을 주관하는 기본 원리가 있고, 그 원리는 인간이 충분히 이해/설명이 가능하다는 믿음이 자연과학 (특히 물리학)에 존재해 왔다. 이러한 흐름에 힘입어, 기독교인 자연과학자들은 하나님이 우주에 단순한 원리를 적용하셨으며, 그 원리를 인간이 이해할 수 있는 축복을 허락하셨다고 믿는 경향이 있다. (Swinburne, 1979; Craig & Sinnott-Armstrong, 2003) 하지만 오늘날 자연과학의 첨단 문제에서 인간이 그 동작 원리를 100% 파악하기 어려운 인공지능을 필요로 한다는 것은, 결국 자연의 동작 원리를 인간이 100% 이해할 수는 없다는 것을 다시금 일깨워주는 것 같다. 물론 인간이 자연의 동작 원리를 어느 정도는 파악할 수 있다는 점은 분명하지만, 다시금 욱기 38~41장을 거치고 난 욱의 고백에 주목할 필요가 있는 것 같다. “주께서는 못 하실 일이 없사오며 무슨 계획이든지 못 이루실 것이 없는 줄 아오니 / 무지한 말로 이치를 가리는 자가 누구니이까 나는 깨닫지도 못한 일을 말하였고 스스로 알 수도 없고 헤아리기도 어려운 일을 말하였나이다” (욱42:2~3)

두 번째로, 인공지능이 눈부신 발전을 거듭하는 앞으로의 시대에는 인간 존재의 존엄성을 인간의 지적 능력에서만 찾기 어려우리란 점이다. 많은 과학철학자들이 단일 기계의 지적 능력이 인류 전체의 지적 능력을 능가하는 ‘기술적 특이점’(technological singularity)가 수십 년 이내에 찾아올 가능성이 있으며, 이 사건이 일어난 후부터는 인류의 과학기술문명의 발전이 오로지 기계 스스로의 힘에 의해 이루어질 것이라고 예측하고 있다. (Shanahan, 2015) 이미 2020년 현재에도 대부분의 인간은 육체적인 능력이나 (계산적인) 지적 능력에서 기계의 능력에 한참 못 미친다. 만약 인간의 가치를 효용성으로 재단한다면, 대부분의 인간의 가치는 기계보다 못하게 되는 것인가? 인문학적인 관점으로는 효용성 이외의 인간의 가치를 추구할 수 있겠으나, 인공지능이 발전하여 인간의 소위 인문학적인 지적 소양을 인공지능이 갖추게 되었을 때에도 인간이 여전히 기계보다 우월하다고 할 수 있을까? 기독교인으로서 우리는 인간 존재의 존엄성의 근본은 하나님으로부터 부여된다는 것을 알고 있다. (창1:27~28) 이러한 인간의 존엄성을 기독교인인 우리가 먼저 삶 속에서 확실히 붙들고, 더 나아가 신앙 공동체 밖의

사회에서도 추구해 갈 수 있도록 노력할 필요가 있다.

## 참고문헌

- 김은. (2016). 『인더스트리 4.0의 연혁, 동향과 방향 전망』. KIET 산업경제.
- Allen-Zhu, Z., Li, Y., & Song, Z. (2019). *A convergence theory for deep learning via over-parameterization*. In International Conference on Machine Learning (pp. 242-252). PMLR.
- Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
- Berlanstein, L.R. (Ed.). (2003). *The Industrial Revolution and Work in Nineteenth Century Europe*. Routledge.
- Carrasco Kind, M., & Brunner, R. J. (2014). *SOM z: photometric redshift PDFs with self-organizing maps and random atlas*. Monthly Notices of the Royal Astronomical Society, 438(4), 3409-3421.
- Chen, C.S. (2009). *Atheism and the Assumptions of Science and Religion*.
- Craig, W.L., & Sinnott-Armstrong, W. (2003). *God?: A Debate between a Christian and an Atheist*. Oxford University Press.
- Dieleman, S., Willett, K.W., & Dambre, J. (2015). *Rotation-invariant convolutional neural networks for galaxy morphology prediction*. Monthly notices of the royal astronomical society, 450(2), 1441-1459.
- Engelman, R. (2015). *The second industrial revolution, 1870-1914*. US History Scene, 10.
- Fukushima, Kunihiko (1979). 『位置ずれに影響されないパターン認識機構の神経回路のモデル --- ネオコグニトロン ---』. Trans. IECE (in Japanese). J62-A (10): 658 - 665.
- Hebb, D.O. (1949). *The organization of behavior: a neuropsychological theory*. J. Wiley; Chapman & Hall.
- Heilbron, J.L. (Ed.). (2003). *The Oxford companion to the history of modern science*. Oxford University Press.
- Hermann, M., Pentek, T., & Otto, B. (2016). *Design principles for industrie 4.0 scenarios*. In 2016 49th Hawaii international conference on system sciences (HICSS) (pp. 3928-3937). IEEE.
- Hopfield, J.J. (1982). *Neural networks and physical systems with emergent collective computational abilities*. Proceedings of the National Academy of Sciences of the USA, vol. 79 no. 8 pp. 2554 - 2558.
- Hoyle, B. (2016). *Measuring photometric redshifts using galaxy images and Deep Neural Networks*. Astronomy and Computing, 16, 34-40.
- Hume, D. (1739). *A Treatise Upon Human Nature*.
- Ioffe, S., & Szegedy, C. (2015). *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. arXiv preprint arXiv:1502.03167.
- Ivakhnenko, Alexey (1971). *Polynomial theory of complex systems*. IEEE Transactions on

- Systems, Man and Cybernetics. SMC-1 (4): 364 - 378.
- Liu, K., Ok, K., Vega-Brown, W., & Roy, N. (2018). *Deep inference for covariance estimation: Learning gaussian noise models for state estimation*. In 2018 IEEE International Conference on Robotics and Automation (ICRA) (pp. 1436-1443). IEEE.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York: springer.
- Janssen, D. *What is the Digital Revolution?*-Definition from Techopedia.
- Jo, Y., & Kim, J.H. (2019). *Machine-assisted semi-simulation model (MSSM): estimating galactic baryonic properties from their dark matter using a machine trained on hydrodynamic simulations*. Monthly Notices of the Royal Astronomical Society, 489(3), 3565-3581.
- Kim, T., Park, E., Lee, H., Moon, Y. J., Bae, S.H., Lim, D., ... & Cho, K.S. (2019). *Solar farside magnetograms from deep learning analysis of STEREO/EUVI data*. Nature Astronomy, 3(5), 397-400.
- McCulloch, W.S., & Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity*. The bulletin of mathematical biophysics, 5(4), 115-133.
- New York Times, July 8th, 1958, "NEW NAVY DEVICE LEARNS BY DOING; Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser"
- Pan, S., Liu, M., Forero-Romero, J., Sabiu, C. G., Li, Z., Miao, H., & Li, X.D. (2020). *Cosmological parameter estimation from large-scale structure deep learning*. Science China Physics, Mechanics & Astronomy, 63(11), 1-15.
- Paracelsus, T. (1537). *De natura rerum*.
- Poole, D., Mackworth, A., & Goebel, R. (1998). *Computational Intelligence*.
- Reddi, S.J., Kale, S., & Kumar, S. (2019). *On the convergence of adam and beyond*. arXiv preprint arXiv:1904.09237.
- Russell, S.J., & Norvig, P. (2003). *Artificial Intelligence A Modern Approach*. Pearson Education.
- Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., ... & Penedones, H. (2020). *Improved protein structure prediction using potentials from deep learning*. Nature, 577(7792), 706-710.
- Shanahan, M. (2015). *The technological singularity*. MIT press.
- Shelley, M.W. (1818). *Frankenstein, or the modern Prometheus*.
- Springenberg, J.T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2014). *Striving for simplicity: The all convolutional net*. arXiv preprint arXiv:1412.6806.
- Swinburne, R. (1979). *The existence of God*. Oxford University Press.
- Wang, J., Yu, F., Chen, X., & Zhao, L. (2019). *Admm for efficient deep learning with global convergence*. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 111-119).
- Wei, G.W. (2019). *Protein structure prediction beyond AlphaFold*. Nature Machine Intelligence, 1(8), 336-337.
- Yi, D., Ahn, J., & Ji, S. (2020). *An Effective Optimization Method for Machine Learning Based on ADAM*. Applied Sciences, 10(3), 1073.

- Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., & Lipson, H. (2015). *Understanding neural networks through deep visualization*. arXiv preprint arXiv:1506.06579.
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). *Learning deep features for discriminative localization*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2921-2929).

논 찬 문		기독교학문연구회	
발표논문 제목	인공지능과 자연의 이해	발표자	홍성욱
		논찬자 (소속)	이경건 (서울대 박사수료)

해당 논문은 인공지능이 자연에 대한 인간의 이해에 얼마나 도움을 줄 수 있는가를 주로 다루고 있습니다. 저자께서 다양한 과학 분야에서 인공지능경망이 활용되고 있는 상황을 소개하는 부분은 폭넓은 영역 지식(domain knowledge)에 기반할 뿐 아니라 일반 독자들을 배려하여 가독성이 높은 것 같습니다. 논문의 막바지에서는 기독교인들이 얻을 수 있는 몇 가지 함의를 제공하는 듯합니다. 전반적으로 보아, 해당 논문의 대요는 대다수의 독자들에게 폭넓게 받아들여 질 수 있으리라고 생각합니다.

다만, 논지의 폭이 다소 제한되어 있다는 점, 글의 전반적인 내용과 논의에서 기독교적 학문 혹은 기독교 세계관과의 연관성이 다소 명확하지 않다는 점, 그리고 해당 논문에서 제기하고자 하는 문제의식이 과연 새로운가 하는 점 등을 짚어볼 수 있겠습니다.

우선, 제한된 논지의 측면입니다. 첫째, 저자께서도 주지하시리라 생각합니다만 인공지능, 인공지능경망, 기계 학습(머신 러닝)을 동일시하는 것은 무리가 있습니다. 인공지능경망의 학습은 인공지능을 구현하는 가장 유망한 방법론임이 분명하나, 그것이 곧 인공지능이라고 하기는 어렵습니다. 인공지능의 정의 자체가 철학적으로 매우 중대한 쟁점임에도 불구하고 이를 과도하게 협의로 사용한 점이 적잖이 아쉽습니다. 둘째로, 인공지능과 자연과학의 어떠한을 논하는 부분에서도 논지가 제한되어 있는 듯합니다. 물론, 이에 관하여 예측력(설명력), 통계적 해석가능성, 과학 이론 발달에의 기여 가능성이라는 세 기준을 세우신 것은 어느 정도 타당합니다. 다만 ‘과학적 방법론’ 자체가 무엇인지, 통계적 모델의 간명성(parsimony) 원리가(e.g. 오컴의 면도날) 인공지능경망에서 점차 쇠퇴하고 있는 현상의 의미가 무엇인지 등에 대하여 정리가 되었다면 더욱 좋았을 것입니다.

다음으로, 글의 서론과 결론 부분에서는 기독교적인 무언가를 암시하고 있는 듯하나 그에 대하여 보다 학술적이고 엄밀한 주장이 명확히 드러나지 않는 것 같습니다.

마지막으로, 해당 논문의 내용(인공지능이 과학 연구에 사용되고 있는 현실, 그에 대한 우려, 기독교적 관점에서의 질문)이 독자들에게 새로운 무언가를 제공하는가에 대하여 약간의 주저가 생깁니다. 제가 알기로 기독교학문연구회에서만도 이미 그러한 기초 하에 학술대회가 열린 적이 있으며, 이에 준하는 발표들 또한 수차례 있었습니다. 저자께서는 어렵고 복잡할 수 있는 인공지능 관련 내용들을 일반 독자들도 수월하게 이해할 만큼 일목요연하게 정리하셨으며, 이에 대하여 많은 정성을 기울이셨으리라 생각합니다. 하지만 해당 논문이 학술대회에서 발표된다는 점에 있어서는 보다 참신성과 논의의 진전에 초점을 두셨더라면 더욱 좋았으리라 생각합니다. 예컨대, 인공지능에 대한 기독교세계관 그룹의 논의는 이제 (1) 인공지능의 정의와 인간론에 대한 기독교 세계관적 함의, (2) 자연과학의 정의와 기독교세계관적 함의, 그리고 (3) 이 양자가 교차하는 지점에 대한 각론으로 발전하여야 할 상황입니다. 훌륭한 개론을 제공해주신 저자께서 앞으로 그러한 기여를 해주신다면 기독교세계관 학계가 큰 도움을 얻으리라 생각합니다.

제37회 기독교학문연구회 연차학술대회	국립목포대학교
----------------------	---------