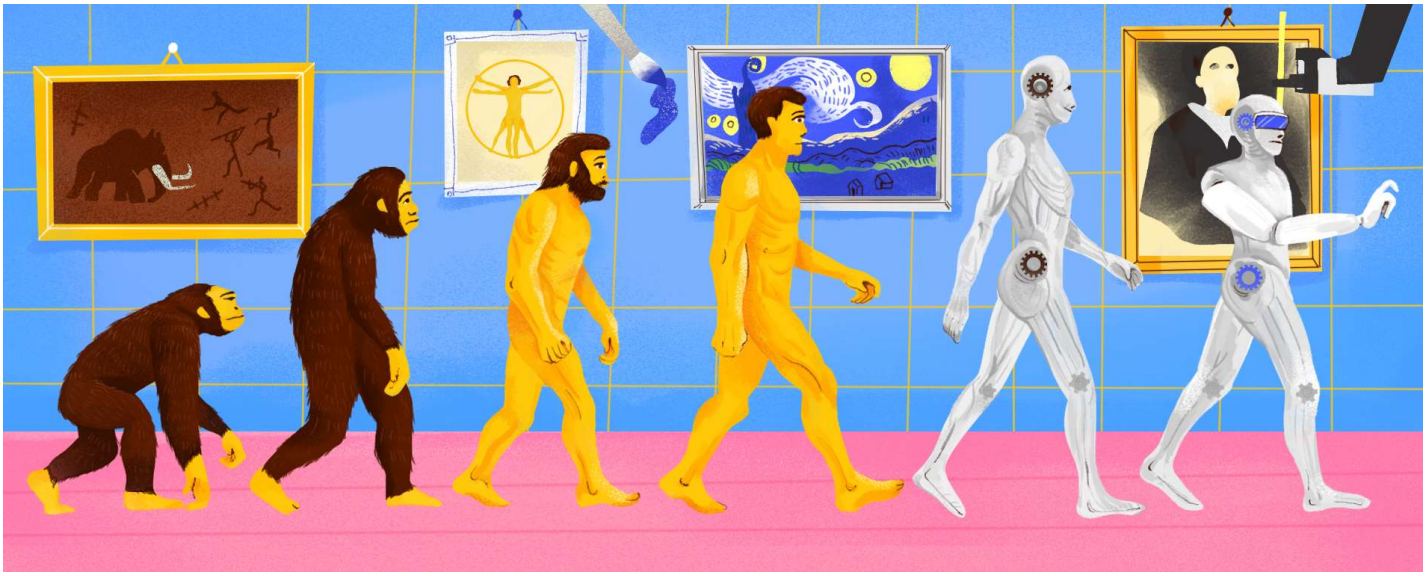
과학철학

인공지능과 포스트휴머니즘 – 초학제의 뉴 호라이즌

2020년 4월 14일

이중원



TRP-19, ‘인공지능과 포스트휴머니즘’

고등과학원에서는 과학연구의 지평을 확대하고, 학문 간 협력을 통해 새로운 학문적 아젠더를 개발하려는 목표를 갖고, 해마다 초학제 연구프로그램Transdisciplinary Research Program, TRP을 운영해 오고 있다. 학제 간 교류와 소통을 통해 특정 전문분야 (혹은 학제) 내에서의 사유의 한계를 극복하고자 하는 일종의 융합연구 프로그램이라고 할 수 있다. 2019년에는 이 연구프로그램의 일환으로 <2019 올해의 주제연구단>이 출범하여 ‘인공지능과 포스트휴머니즘’을 주제로 다양한 학술 활동을 추진하였다. 이는 전문가 상호 간에 통섭적인 학술교류는 물론, 전문가와 일반인이 만나는 소통의 장이라는 의미를 지닌다. 이글에서는 초학제 연구의 주제로 ‘인공지능과 포스트휴머니즘’이 선택된 배경을 설명하고, 초학제 프로그램에서 특별히 강조되었던 연구 주제들이 미래에도 지속되어야 할 연구 테마라는 관점에서 소개하며, 이러한 연구 테마들이 고등과학원이 추진하는 초학제 연구의 새로운 지평을 여는 데 어떤 의의가 있는지 시사점을 제시해보고자 한다.

‘인공지능과 포스트휴머니즘’이라는 주제는 인공지능 기술의 급속한 발전으로 생활양식과 사회관계가 급격히 변화하고, 그에 수반하는 윤리적·법적·사회적 쟁점들이 새롭게 부상하며, 급기야 휴머니즘 및 인간의 정체성에도 많은 변화가 나타날 것이라는 문제의식에서 출발한 것이다. 초학제 연구의 취지에 맞도록 다음과 같은 연구 목표를 설정하였다. 첫째, 21세기 과학기술문명을 선도하고 있는 인공지능 기술을 중심으로 그것의 발전 배경과 현황 및 전망을 기술적 측면에서 구체적으로 살펴보는 것이다. 둘째, 이러한 기술의 발전이 가져올 변화들, 곧 인간과 인공지능 간 새로운

관계 설정, 사회 패러다임의 변화 그리고 이에 수반하는 윤리적·법적·사회적 쟁점들과 같은 인간학적 문제들을 선제적으로 예측하고 분석하는 것이다. 셋째, 인공지능의 인격체로의 발전 가능성, 인공지능을 활용한 인간의 능력증강과 그에 따른 인간 정체성의 변화, 그리고 기계의 인간화 경향과 인간의 기계화 경향이 교차하는 미래의 포스트휴먼 환경에서 궁극적으로 인간과 인공지능이 함께 공존할 수 있는 적합한 담론으로서 포스트휴머니즘을 모색하는 것이다. 이러한 주제 연구를 위해서는 학제 간 연구 방식, 특히 과학기술 분야와 인문사회 분야가 한자리에 모여 서로 소통하고 융합하는 통섭적 접근방식이 필수적이다.

왜 '인공지능과 포스트휴머니즘'에 주목하는가?

인류 역사에서 17~18세기의 근대혁명은 근대적 개인, 근대적 사회 그리고 휴머니즘(개인의 주체성과 존엄성을 중시하는 인본주의 사상)을 탄생시켰다. 그리고 근대철학은 개인을 주체적 존재로 보고, 그러한 개인에게 인간 본성의 핵심요소라 할 수 있는 이성, 감성, 도덕성, 가치, 자의식, 자유의지 등이 존재함을 강조하였다. 이를 토대로 인간과 인간이 아닌 다른 것들을 차별화하여 그 경계를 명확히 하고, 다른 모든 것들은 주체인 인간을 중심으로 그 주위에 마주하는 객체로 보는 인간중심주의가 탄생했다. 이에 따르면 자연의 사물들은 인간에 의해 그 특성이 감지되고 인지될 때, 비로소 존재적 의미와 가치를 부여받게 된다. 모든 존재가 본래의 고유한 가치를 상실한 채, 주체인 인간에 의해 규정당하고 대상화되는 것이다. 근·현대 과학기술문명 역시 이러한 인간중심주의 기반 위에 서 있다. 인간은 객체인 자연을 탐구하는 주체로서 존재하고, 자연의 모든 정보는 인간이 설계한 관측 장치나 실험도구에 의해 인간이 감지할 수 있는 형태로 수집・분석되며, 자연의 모든 법칙과 현상들은 인간이 만들어 낸 언어 및 관념체계에 의해 규정・해석되도록 구축되었다. 결국 인간과 신, 인간과 인간, 인간과 자연, 인간과 기계의 모든 관계가 인간중심적 관점에서 언급되는 시대가 되었다.

하지만 21세기가 되면서 이러한 인간중심주의에 변화가 일어나고 있다. 우선 인간과 동물의 관계는 차치하고 인간과 기계의 탈경계화를 가속하는 과학기술이 빠른 속도로 발전하고 있다. 과거 인간과 기계는 서로 이질적인 존재로 그 경계가 명확했지만, 21세기에는 인간의 기계화 경향과 기계의 인간화 경향이 두드러지면서 인간과 기계 간 경계가

약화되고 있다. 예를 들어 신체 일부를 인공장기나 로봇 팔·다리로 대체하는 신체변형기술, 자유롭게 유전자를 조작하는 생명편집기술, 인공세포나 인공혈액을 만드는 나노기술, 사이보그 기술의 발전은 인간의 능력증강Human Enhancement을 가능하게 하는 인간의 기계화 경향의 좋은 사례들이다.

반면 인간의 감정을 표현하고, 인간처럼 생각하고 말하며 행동하는 휴머노이드 로봇의 등장은 기계의 인간화 경향을 잘 보여준다. 특히 이는 그동안 인간에게 고유한 것으로 인식됐던 능력들(감성, 이성, 도덕성 등)이 (인간과 동일하진 않지만) 기계에서도 구현 가능한 인공지능 기술이 발전하면서 본격화되었다고 할 수 있다. 아직은 인공지능이 특정 영역에서만 인간의 능력을 뛰어넘을 뿐 인간처럼 다양한 영역에서의 멀티 능력을 갖추고 있지 못하고 있고, 로봇 역시 언어나 감정 표현이 아직은 서툴고 행동 또한 인간처럼 유연하지도 민첩하지도 섬세하지도 못하다는 한계는 분명히 존재하지만, 시간이 흐를수록 이러한 한계들은 점차 극복될 것이다. 먼 미래의 특이점에 접근할수록 초지능을 지닌 새로운 종으로서 '로보 사피엔스'의 등장 가능성도 배제할 수 없을 것이다.

"

인간과 기계의 탈경계화는 기계를 바라보는 인간의 시각을 바꾸고, 인간과 기계의 관계를 새롭게 정립하며 궁극적으로 휴머니즘에 변화를 야기할 것으로 예상된다.

"

이러한 인간과 기계의 탈경계화는 기계를 바라보는 인간의 시각을 바꾸고, 인간과 기계의 관계를 새롭게 정립하며 궁극적으로 휴머니즘에 변화를 야기할 것으로 예상된다. 지금까지 인간은 기계를 그 존재적 특성과 무관하게 인간중심적 관점에서 단순한 도구로 간주해 왔다. 하지만 인공지능에서 보듯이 기계는 인간의 어떤 목적을 위한 수단이 아니라, 그 자체가 어떤 목적을 갖고 세계를 구성하는 실존적 존재자로 변화하고 있다. 단순히 인간 사회를 떠받쳐주는 물리적 기반이 아니라, 인간과 복잡하게 연결된 관계망 속에서 사회를 구성하는 하나의 행위자actor 혹은 agency로 거듭나고 있다. 한마디로 인공지능 시대에 기계는 더이상 인간의 외부에 존재하는 객체가 아니라, 인간의 몸과 마음의 일부로 주체화될 수 있는 존재인 것이다. 아직은 일부의 현상이지만 섹스 로봇과 결혼한다거나, 인공지능 로봇에게 시민권을 부여하는 사건들은 이러한 단면을 잘 보여 준다.

이렇듯 인간의 기계화 경향과 기계의 인간화 경향이 상호 교차하면서, 인간과 기계의 탈경계화가 가속화되는 상황 혹은 그러한 상황을 가능하게 하는 조건을 포스트휴먼의 상태라 말할 수 있겠다. 그런 의미에서 21세기는 포스트휴먼 시대다. 이는 포스트휴머니즘이 확실하게 정착된 단계는 아니며, 휴머니즘 시대에서 포스트휴머니즘 시대로 넘어가는 과도기적 단계라고 말할 수 있다. 이러한 시기에는 포스트휴먼을 이끌어가는 과학기술의 발달이 인간과 기계의 관계에 어떤 변화를 일으키는지, 그로 인해 인간 정체성에는 구체적으로 어떤 변화가 나타날지, 나아가 달라진 인간 정체성의 시각에서 미래사회는 앞으로 어떻게 변화할 것인지, 그리고 그러한 사회의 지속가능한 발전을 위해 어떠한 윤리적·법적 규범과 사회 제도적 장치들이 새롭게 필요한지에 관한 선제적 연구가 필요하다. 이는 곧 다가올 미래사회에 대한 초학제적 대응의 의미를 갖는다.

‘인공지능과 포스트휴머니즘’의 주요 연구 테마들

2019년에 추진했던 초학제 연구의 가장 중요한 특징은, 현재 주목받고 있거나 가까운 미래에 중요하게 부각될 미래 인공지능 기술을 중심으로 이의 연구·개발 현황 및 발전 전망을 상세하게 분석하고, 이러한 인공지능 기술의 연구·개발 및 상용화 과정에서 고려해야 할 인문적·사회적 요소들을 함께 성찰함으로써, 두 영역 간 상호보완 작용을 통해 생산적 융합을 지향하도록 설계·운영한 점이다. 실제로 학술 활동에서 세부주제가 정해지면 관련 과학기술 분야의 전문가와 인문사회 분야의 전문가 각 1인씩 발표하고, 종합토론자가 두 이야기를 통합하는 통섭적 접근방식으로 진행됐다. 미래에도 계속되어야 할 학제 간 연구주제로, 많은 사람의 관심과 이목이 집중됐던 몇 가지 연구 테마를 소개하고자 한다.

첫 번째 연구 테마는 ‘설명 가능한 인공지능’Explainable AI, XAI에 관한 것이다. 알파고나 왓슨처럼 심층학습deep learning을 통해 스스로 학습하는 인공지능을 넘어, 앞으로는 인공지능의 판단에 대해 외부에서 설명을 요청하는 경우 그 판단이 어떤 근거 및 절차를 통해 이루어졌는지 스스로 해명하는, 소위 ‘설명 가능한 인공지능’이 차세대 인공지능으로 주목받고 있다. 실제로 이러한 설명 능력은, 예를 들어 자율형 군사 킬러 로봇의 경우 매우 중요한 관건이 될 수 있다. 가령 킬러 로봇이 민간인을 적군으로 착각하여 죽였을 경우 그에 따른 책임 문제가 당연히 뒤따를 텐데, 이때 킬러 로봇이 어떤 근거와 절차로 그런 판단을 내렸는지에 대한 설명이 매우 중요하다. 그래야 그 설명을 토대로 어디서 오류가 발생했는지를 발견할 수 있고 그에 따른 책임 소재를 규명할 수 있기 때문이다. 아직 까지는 인간이 아닌 킬러 로봇과 같은 인공 행위자에 책임을 부여할 수는 없기에 설명의 문제는 사회적으로 매우 중요하다.

그림1 국제학술토론회 <Beyond Humanism: AI, Information and Posthumanism>

2019 올해의 주제연구단 ‘인공지능과 포스트휴머니즘’

그런데 이러한 인공지능에서 설명 과정이 성공적으로 작동하려면, 어떤 설명이 편견이나 왜곡 없는 합당한 설명인지를 결정하는 알고리즘도 중요하지만, 그에 앞서 인간이 요청한 설명이라는 개념 자체가 정확히 무엇을 의미하는지에 대한 이해가 인공지능에 반영돼 있어야 한다. 그래야 인공지능이 설명에의 요청을 정확히 간파하고 인간이 이해할 수

있는 방식으로 설명을 제공할 수 있기 때문이다. 이는 매우 개념적이고 인식적인 문제로서, 언어학이나 철학과 같은 인문학과의 학제 간 연구를 필요로 한다.

두 번째 연구 테마는 인공지능과 인간의 뇌에서의 학습에 관한 것이다. 알파고를 포함한 현재의 인공지능은 인간의 신경망 모델에 기반한 심층학습을 통해 진화를 거듭하고 있다. 초창기 인공지능(한 예로 알파고-리 버전)이 정답을 알려주며 주어진 빅데이터를 활용해 정답을 찾아 나가도록 하는 지도학습에 기반했다면, 최근의 인공지능(한 예로 알파고-제로 버전)에서는 정답을 알려주지 않고 비슷한 데이터를 군집화하여 최상의 미래를 예측하도록 하는 비지도 학습이 일반화돼 있다. 한층 진화된 자기 주도적 학습인 셈이다.

하지만 인공지능에서 이러한 학습 과정은 근본적인 한계를 지니고 있다. 인간은 적은 수의 데이터일지라도 대상의 속성을 정확히 파악하고 이를 범주화하여 다른 대상과 쉽게 분별하며, 이렇게 설정된 범주를 개념화하여 뇌에 기억하는 방식으로 대상에 관한 매우 효율적이고 효과적인 학습을 전개한다. 하지만 인공지능에서는 이 작업 과정 자체가 현재로서는 불가능하다. 범주적이고 개념적인 접근 없이, 대상에 관한 다수의 데이터에 의존한 패턴 인식만으로 대상을 분별하고 학습한다. 물론 미래에는 이보다 훨씬 진화하여 인간처럼 범주적 사고를 통한 개념학습이 가능할지도 모른다. 그렇게 되려면 인간의 뇌에서 개념학습이 어떻게 이루어지는지 그 메커니즘을 규명하고, 인간과 인공지능의 학습 과정에서 공통점과 차이점이 무엇인지를 명확히 분석하며, 이에 기반한 새로운 기계학습Machine Learning 모델을 개발하는 것이 필요하다. 이를 위해서는 뇌 과학, 컴퓨터 과학, 교육학 분야 간에 학제 간 융합 연구가 중요하다.

"

초창기 인공지능이 정답을 알려주며 주어진 빅데이터를 활용해 정답을 찾아 나가도록 하는 지도학습에 기반했다면,

최근의 인공지능에서는 정답을 알려주지 않고 비슷한 데이터를 군집화하여 최상의 미래를 예측하도록 하는 비지도 학습이 일반화돼 있다.

"

세 번째 연구 테마는 인공지능과 인간의 언어에 관한 것이다. 최근에 등장한 인공지능 스피커, 시리나 빅스비 같은 대화 앱, 파파고 등에서 보듯이 기계학습 및 빅데이터를 활용해 인공지능이 인간과 대화하거나 자동 번역 및 통역을 수행하는 의사소통 기술이 급속도로 발전하고 있다. 이러한 속도라면 인공지능이 인간처럼 인간과 자연스럽게 대화할 날도 멀지 않은 것 같다. 하지만 인공지능이 지금처럼 인간의 언어 활동을 돕는 보조자가 아니라, 인간처럼 인간의 언어를 충실히 이해하고 인간과 자연스럽게 대화를 나누는, 또 하나의 언어 행위자가 되는 일은 결코 간단한 일이 아니

다. 인간의 일상언어를 자연스럽게 이해하려면, 언어로 표현된 기호가 어떤 맥락에서 어떤 의미로 쓰였는지를 정확히 파악해야 하는데, 현재의 인공지능에서는 기호의 의미 파악과 맥락에 따른 의미 변화 모두를 이해하기가 매우 어렵다.

이런 일이 가능하려면 많은 도전적인 문제들이 해결돼야 한다. 그 가운데 가장 중요한 관건이 되는 문제는 아마도 인간의 뇌에서 언어 활동, 특히 의미 설정이 어떻게 이루어지는지 그 과정을 탐구하고 메커니즘을 규명하는 것이다. 그런 다음 이를 토대로 현재 인공지능의 언어처리 과정이 인간의 언어처리 과정과 어떤 면에서 유사하고 어떤 면에서 다른지를 명확히 밝히는 것이다. 그리고 이런 비교 위에서 인공지능이 인간과 자연스럽게 일상언어로 대화하는데 필요한 부분이 무엇인지 언어(학)적 측면에서 그리고 기술적 측면에서 고찰하는 것이다. 최종적으로는 이를 바탕으로 인간을 모사한 자연언어 처리 프로그램을 인공지능에 설계하는 것이다. 한편 이렇게 인간과 자연스럽게 대화할 수 있는 인공지능이 등장한다면, 인간 사회에 어떤 일이 벌어질까? 아마도 다양한 영향을 미칠 것이므로 선제적 차원에서 그 사회적 영향에 대한 학제 간 논의 역시 필요하다.

네 번째 연구 테마는 인공지능과 예술에 관한 것이다. 현재 예술 창작활동을 하는 수많은 인공지능이 존재한다. 화가로 활동하는 알고리즘 가운데, 구글의 딥드림Deep Dream은 동일 구조가 비슷하게 반복되는 프랙털fractal 형태의 그림을 그리는데, 작품 29점이 2016년 샌프란시스코 경매에서 9만 7000달러에 판매되었다. 프랑스가 개발한 인공지능 오비어스Obvious가 그린 초상화 '에드몽 드 벨라미'는 2018년 뉴욕 경매에서 43만 2500달러에 낙찰되었다. 기존의 그림을 유명 예술가의 화풍으로 변형시켜 새로운 그림으로 만드는 딥포저Deep Forger, 램브란트의 화풍을 그대로 재현해 내는 넥스트 램브란트 등도 화가로 활동하고 있다. 또한 음악을 작곡하는 소니의 플로머신, 시를 쓰는 중국의 샤오이스, 소설을 쓰는 미국의 셸리처럼 음악이나 문학 분야에서 창작활동을 하는 인공지능도 다수 존재한다. 가히 인공지능 예술창작시대, 혹은 인공 창의 시대라 언급할 만하다. 그동안 인간만의 배타적인 창작 영역으로 인정돼왔던 예술조차 인공지능에 의해 구현 가능한 시대가 된 것이다.

그림2 <Edmond de Belamy>. 그림 하단에 알고리즘이 적혀 있으며, 제작 과정은 다음 동영상에서 확인 가능하다.

Obvious

그렇다면 다음과 같은 질문들이 제기될 수 있다. 인공지능도 인간처럼 예술가가 될 수 있는가, 만약 예술가가 될 수 있다면 핵심 요건인 창의성이 인공지능에서 어떻게 구현될 것인가, 만약 예술가가 아니라면 인공지능이 만든 창작품은 예술작품이 될 수 있는가 등등. 이 질문들은 다시 예술이란 무엇인가, 창의성이란 무엇인가라는 원초적 질문으로 되돌아가, 이에 관한 많은 논의가 필용하다.

초학제 연구의 새로운 지평New Horizon 찾기

초학제 연구프로그램은 그 취지와 목적이 매우 의미 있고, 21세기 인공지능으로 대표되는 포스트휴먼 시대에 매우 중요하고 필요하다. 그래서 초학제 연구프로그램이 앞으로 더욱 활성화되기 위해 지금 무엇이 필요한지 논의하는 것은 매우 큰 의미가 있다고 생각한다. 특히 초학제 연구프로그램이 고등과학원에서 운영하는 만큼, 그 연구사업들이 고등과학원의 취지와 역할에 걸맞은 융합연구로서 성공적으로 운영되고 있는가에 대한 평가가 매우 중요하다고 생각한다. 초학제 연구프로그램의 주목표는 학문 간 협력을 통해 새로운 창의적 학문 아젠더를 발굴하여 과학연구의 융합적 지평을 확대하는 것이다.

이 어려운 목표를 성공적으로 달성하기 위해선 두 단계의 과정이 필요하다. 하나는 창의적인 학문적 아젠더를 발굴하기 위해 이질적인 학문 분야가 서로 만나 협력하여 융합연구를 위한 우호적 환경을 조성하는 과정이다. 이 경우 다양한 학문 분야가 만나 폭넓게 대화하는 것이 의미 있다. 다른 하나는 아젠더에 따라 실질적으로 생산적인 융합연구를 수행하는 과정으로, 융합연구의 보다 구체적이고 뚜렷한 목표 설정과 통합적인 연구 전략이 중요하다. 그동안 우리 사회에선 수많은 다양한 융합연구들이 추진됐지만, 엄밀히 말해 많은 경우 학제 간 (느슨한) 협력에 바탕 한 느슨한 융합이 주를 이루고, 학제 간 통합에 기반한 강력한 융합연구는 매우 드물다고 말할 수 있다. 초학제 연구프로그램의 경우도 이와 비슷한 상황이라 할 수 있다. 따라서 고등과학원에 보다 의미 있고 미래지향적인 창의적 융합 연구가 초학제 연구프로그램을 통해 이루어질 수 있도록, 초학제 연구프로그램의 전략을 일부 수정·보완하는 것이 필요해 보인다. 이와 관련한 몇 가지 시사점을 던져보고자 한다.

그림3 제1회 워크숍 <설명가능 인공지능의 발전과 그 사회적 함의>

첫째, 학제 간 느슨한 협력보다는 초학제라는 단어가 의미하는 바 그대로, 학제 간 통합에 바탕 한 실질적인 융합 연구가 활성화될 수 있도록 초학제 연구프로그램의 주제를 선별하는 것이 필요해 보인다. 가령 앞서 2019년에 수행했던 '인공지능과 포스트휴머니즘'의 주요 연구 테마들을 살펴보면, 대부분이 학제 간 통합에 기반하지 않는다면 연구 성과가 제대로 나올 수 없는 연구주제들이다. 또 다른 연구주제의 사례로 최근 정부가 추진해온 휴먼플러스 융합연구개발 첼린지 사업의 연구 테마들을 생각해 볼 수 있다. 가령 인공지능을 활용하여 인간 개개인의 지능이나 신체 또는 오감의 기능을 업그레이드할 수 있는 인간능력증강 기술을 개발하는 연구들의 경우, 학제 간 통합에 바탕 한 전형적인 융합 연구의 주제들로 볼 수 있다. 사실 이 연구주제들도 '인공지능과 포스트휴머니즘'이라는 대주제 안에 포함시킬 수 있다.

둘째, 초학제 연구프로그램이 학제 간 통합에 바탕 한 실질적인 융합 연구를 위한 주제를 선별한다고 하더라도, 그 연구를 직접 수행할 수 있는 조건을 현재로서는 갖추고 있지 않은 상태이므로, 현 단계에서는 그 연구들이 활성화될 수 있는 촉매 창구 역할을 수행하는 것이 바람직하다고 생각한다. 이는 지금과 같은 방식과 형태를 유지하면서도 충분히 가능하다고 생각한다. 가령 선별된 주제를 놓고 관심 있는 여러 학문 분야가 참여하여 학제 간 통합에 바탕 한 융합 연구를 고민하고 기획할 수 있는 공론의 장을 마련해 주는 것이다. 이러한 공론의 장을 통해 창의적인 다양한 융합 연구가 사전에 기획되도록 지원함으로써, 국내의 다양한 연구자들이 이후 국가로부터 지원을 받을 수 있는 가교 역할을 할 수 있을 것으로 기대된다.