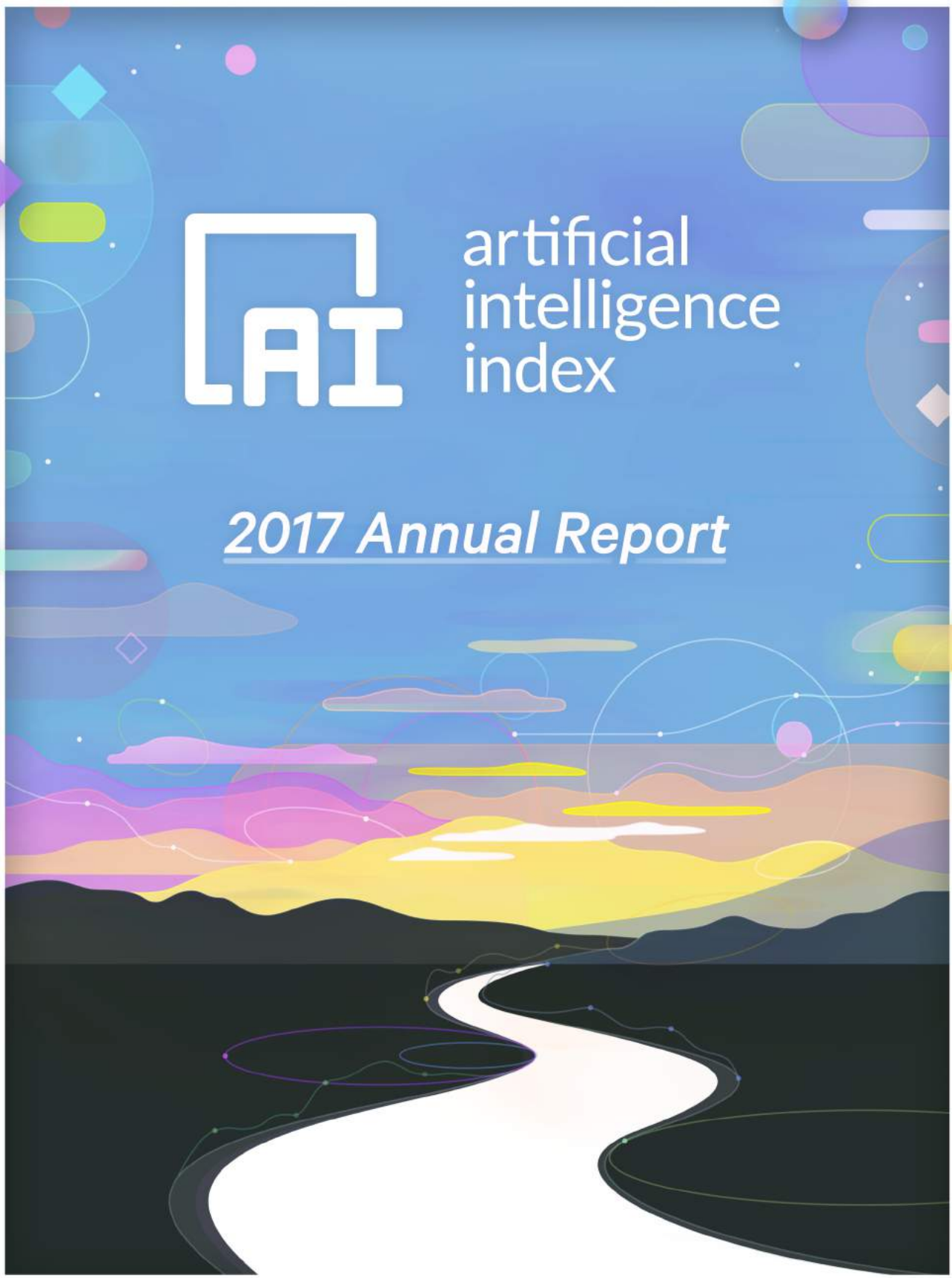
artificial intelligence index

2017 Annual Report



AI 인덱스 2017 연례보고서

요아브 쇼함, 레이몬드 페로, 에릭 브린욜프슨, 잭 클라크, 캘빈 르가식 지음

이상훈, 이동훈 옮김

운영위원회

요아브 쇼함(Yoav Shoham) (위원장)
Stanford University

레이몬드 페로(Raymond Perrault)
SRI International

에릭 브린욜프슨(Erik Brynjolfsson)
MIT

잭 클라크(Jack Clark)
OpenAI

프로젝트 매니저

캘빈 르가식(Calvin LeGassick)

본 보고서는 원문 저자인 AI 인덱스 팀의 허락 하에 번역한 것이다. AI 인덱스 팀은 번역상의 오류에 대해 책임지지 않는다.

보고서 출처 표기법

Yoav Shoham, Raymond Perrault, Erik Brynjolfsson, Jack Clark, and Calvin LeGassick, "The AI Index 2017 Annual Report", AI Index Committee of the One Hundred Year Study on Artificial Intelligence (AI100), Stanford University, Stanford, CA, November 2017. Korean translation by Sanghun Lee, Donghun Lee. Doc: http://cdn.aiindex.org/AI-Index-2017-Korean-Translation.pdf. Accessed: April 26, 2018.

(c) 2017 스탠포드 대학교, 『AI 인덱스 2017 연례 보고서』는 크리에이티브 커먼즈 라이선스 중 저작자표시-변경금지 4.0 국제(CC BY-ND 4.0)에 따라 이용할 수 있습니다.

https://creativecommons.org/licenses/by-nd/4.0/deed.ko

(c) 2017 by Stanford University, "The AI Index 2017 Annual Report" is made available under a Creative Commons Attribution-NoDerivatives 4.0 License (International)

https://creativecommons.org/licenses/by-nd/4.0/deed.en

목차

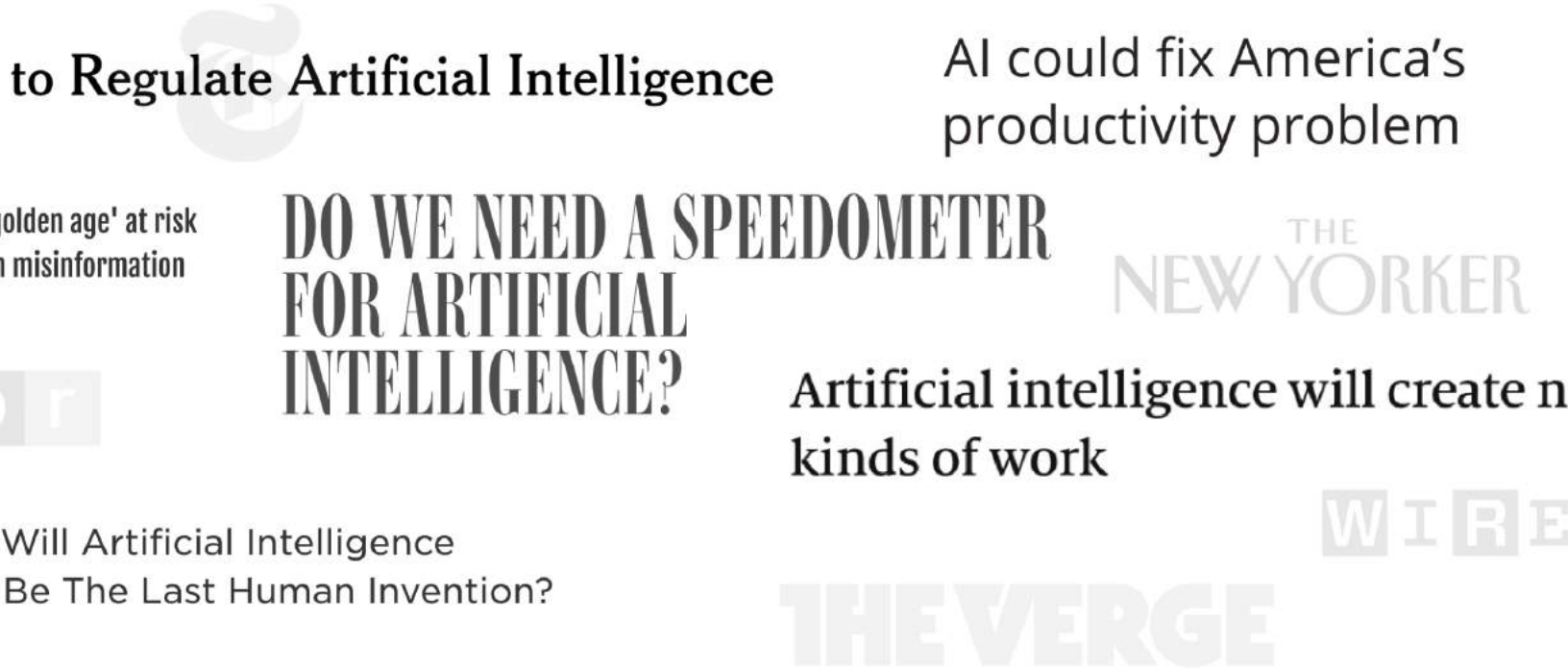
AI 인덱스 2017 연례 보고서 소개 7
개요 9
활동량 11
학계 11
출간된 논문 11
수강 신청 13
컨퍼런스 참석자 16
산업계 18
인공지능 관련 스타트업 18
인공지능 관련 스타트업 투자 19
일자리 20
로봇 수입량 23
오픈소스 소프트웨어 25
깃허브(GitHub) 프로젝트 통계 25
대중적 관심 27
언론 보도 동향 27
기술적 성능 28
시각(Vision) 28
객체 인식(Object Detection) 28
시각 데이터 문답(Visual Question Answering) 29
자연어 이해(Natural Language Understanding) 30
파싱(Parsing) 30

기계 번역(Machine Translation) 31
질의응답(Question Answering) 32
음성 인식(Speech Recognition) 33
정리 증명(Theorem Proving) 34
충족 가능성 문제 해결(SAT Solving) 35
파생된 지표(Derivative Measures) 36
인공지능 기술과 인간 능력 비교 39
보고서에서 다루지 못한 내용 44
전문가 논평 47
참여하기 75
감사의 말 77
부록 A: 데이터 설명 및 수집 방법 79

AI 인덱스 2017 연례 보고서 소개

인공지능(AI) 기술은 실무자, 산업계 리더, 정책 입안자 및 일반 대중의 관심을 받게 되었으며 세계적인 담론의 중심 주제로 자리 잡았다. 올해 언론 기사에서 다뤄진 의견 및 논의의 다양성은 인공지능 기술이 얼마나 광범위하게 조사, 연구, 적용되고 있는지 보여준다.

하지만 인공지능 기술 분야는 여전히 빠르게 진화하고 있으며 전문가들도 전체 영역을 아우르는 내용을 이해하거나 추적하는데 어려움을 겪고 있다.



인공지능 기술 현황에 대해 추론할 수 있는 관련 데이터가 없다면, 우리는 인공지능 기술과 관련한 대화 및 의사결정을 함에 있어 본질적으로 '장님 비행'을 하는 것과 마찬가지다.

우리는 인공지능 기술과 관련한 대화 및 의사결정을 함에 있어 본질적으로 '장님 비행'을 하는 것과 마찬가지다.

스탠포드 인공지능 100년 연구(*One Hundred Year Study on AI at Stanford University*, AI100)의 한 프로젝트로 출발한 AI 인덱스 프로젝트는 인공지능 관련 활동과 발전을 추적하기 위한 오픈소스 기반의 비영리 프로젝트이다. AI 인덱스의 목적은 인공지능 기술에 대해 데이터를 기반으로 한 정보 기반의 대화를 촉진하는 것이다. 본 보고서는 AI 인덱스의 첫 연례 보고서이며, 다양한 관점에서 인공지능 기술 관련 활동과 발전을 살펴보았다. 이를 위하여 인터넷에 공개된 데이터를 종합하고, 원본 데이터를 더하고, 데이터 시리즈를 조합하여 새로운 지표를 만들었다.

본 보고서를 작성하는데 사용된 모든 데이터는 AI 인덱스 웹사이트(aiindex.org)에 공개되어 있다. 하지만 데이터를 제공하는 것은 시작에 불과하며 진정으로 유용한 정보를 제공하기 위해서 AI 인덱스는 더 큰 커뮤니티의 도움이 필요하다. 궁극적으로 본 보고서는 더 많은 사람들이 추가적인 데이터를 제공하고, 수집한 데이터를 분석하고, 추적하기 원하는 주제를 공유하도록 격려하고자 작성되었다. 본 보고서를 통해 여러분이 AI 인덱스 팀과 자유롭게 소통하고 인공지능에 관한 담론의 기반을 마련하고자 하는 노력에 동참하게 되기를 바란다.

개요

보고서 전반부는 AI 인덱스 팀이 종합한 데이터를 다룬다. 보고서 후반부는 보고서에서 다루지 못한 중요 내용에 대한 논의, 보고서에서 다룬 동향에 대한 전문가 논평, 데이터 수집에 관한 도움을 구하고 인공지능 기술 발전에 대해 소통하고 측정하는데 참여를 촉구하는 내용으로 이루어져 있다.

데이터 섹션(Data Sections)

본 보고서에서 수집한 데이터는 크게 네 가지 부문으로 구분된다.

- 활동량(Volume of Activity)
- 기술적 성능(Technical Performance)
- 파생된 지표(Derivative Measures)
- 인공지능 기술과 인간 능력 비교(Towards Human-Level Performance?)

활동량 지표는 인공지능 기술 분야를 '얼마나 많은'의 관점으로 바라본다. 얼마나 많은 참석자가 인공지능 기술 컨퍼런스에 참가했는지 또는 얼마나 많은 벤처캐피털 투자가 인공지능 기술 스타트업을 대상으로 이루어졌는지를 예로 들 수 있다. 기술적 성능 지표는 인공지능 기술을 '얼마나 잘'의 관점으로 바라본다. 컴퓨터가 이미지를 얼마나 잘 이해하고, 수학 이론을 얼마나 잘 증명하는지 등을 측정한다. 데이터셋을 수집하기 위해 사용한 방법에 관해서는 부록에 상세히 설명되어 있다.

처음 두 데이터 셋의 그래프는 이미 잘 알려진 사실을 재확인 시켜준다. 우상향 형태를 보이며 인공지능 관련 활동의 증가 및 기술의 발전을 보여준다. 파생된 지표에서는 트렌드 사이의 관계를 조사하였고, AI 활성 지수(AI Vibrancy Index)라는 실험적인 지수를 소개했다. AI 활성 지수(AI Vibancy Index)는 학계와 산업계의 트렌드를 결합해 인공지능이 하나의 분야로 얼마나 활성화되어 있는가를 측정한다.

인공지능 시스템의 성능을 평가할 때 가장 자연스러운 방법은 인간의 수행능력과 비교해보는 것이다. 인공지능 기술과 인간 능력 비교 섹션에서는 인공지능 시스템이 인간에 준하거나 능가하는 상당한 발전을 이룬 주목할만한 분야에 대해 간략하게 다루었다. 또한 본 보고서에서는 해당 비교의 어려움과 결과를 해석할 때 염두 해야 할 부분에 대해서도 다루었다.

토의 섹션(Discussion Sections)

데이터를 살펴본 후, 보고서에서 다룬 동향에 대한 논의와 중요하지만 보고서에서는 누락된 내용에 대해 다루었다.

토의 섹션에서는 보고서가 가진 한계에 대해서도 다루었다. 본 보고서는 미국 중심의 데이터에 편향된 모습을 보이며, 잘 정의된 벤치마크만을 추적함으로써 기술의 발전을 과대평가하고 있을 수도 있다. 또한 보고서에서는 데이터를 인구 통계학적으로 세분화하여 분석하고 있지 않으며, 정부 및 기업의 인공지능 분야 연구 개발(R&D) 투자에 대한 자료를 포함하고 있지 않다. 본 보고서에서 다루지 못한 해당 부분은 매우 중요하기 때문에 추후에 발간되는 보고서에서 다룰 것이다. 보고서에서 다루지 못한 내용에 관한 섹션에서는 본 보고서의 한계에 대해 자세히 다루고 있다.

본 보고서의 한계를 통해 알 수 있듯이, AI 인덱스는 언제나 인공지능 분야의 일부만을 보여주게 될 것이다. 그렇기 때문에 다양한 분야의 AI 전문가들의 논평을 통해 이를 보완하고자 하였다. 전문가 논평 섹션에서는 본 보고서에서 다룬 데이터에 대한 생생한 배경 설명과 함께 본문에서 충분히 다루지 못한 데이터에 대한 해석을 제시한다.

마지막으로, 보고서 후반부 참여하기 섹션에서는 더 많은 사람들이 이런 논의에 참여할 수 있는 방법을 다루고 있다. 본 보고서에서 다룬 문제를 해결하고, 본문에서 다루지 못한 내용을 탐구하며, 인공지능 분야의 발전 및 활동을 측정하고 생산적인 진전을 이루어 내기 위해서는 더욱 큰 커뮤니티의 참여와 피드백이 필요하다.

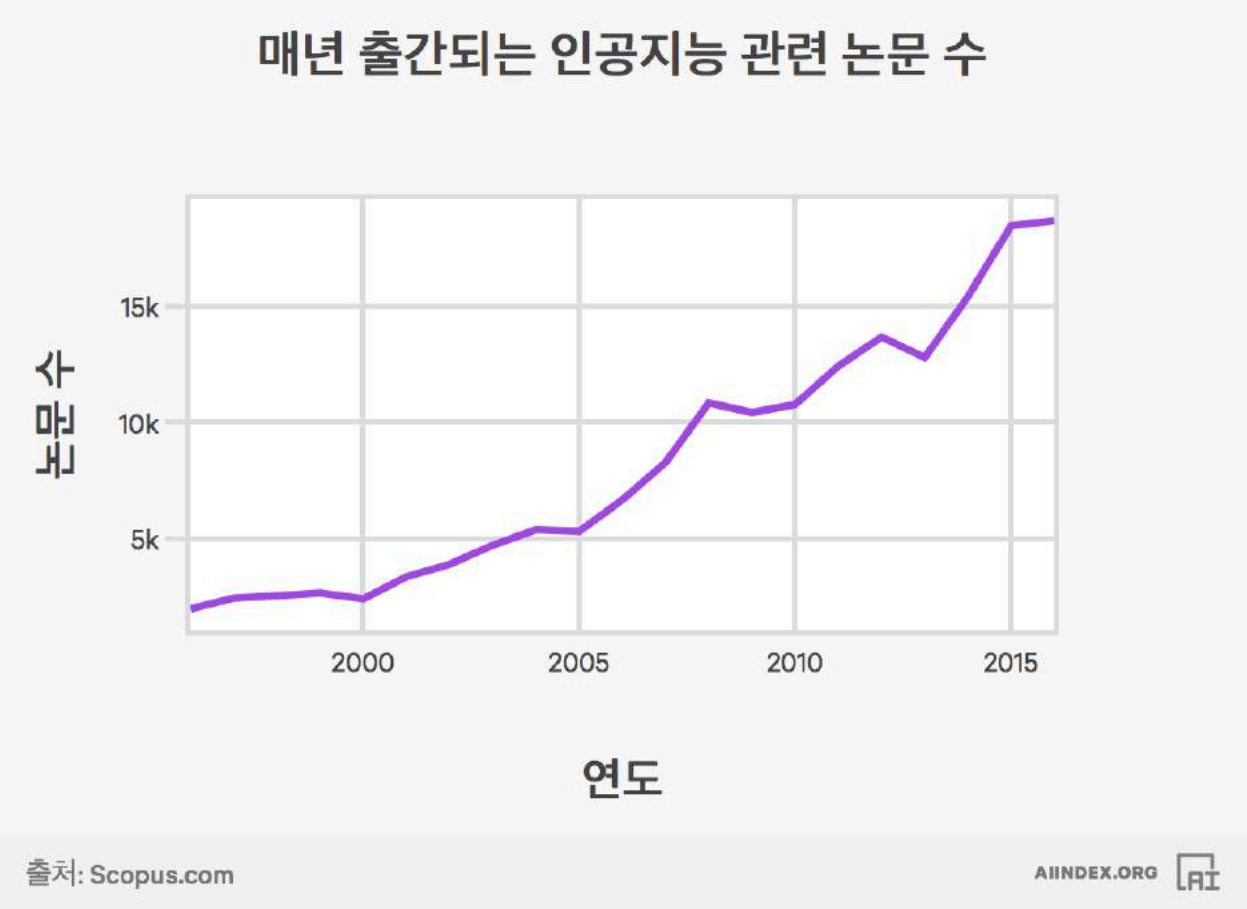
활동량

학계

출간된 논문

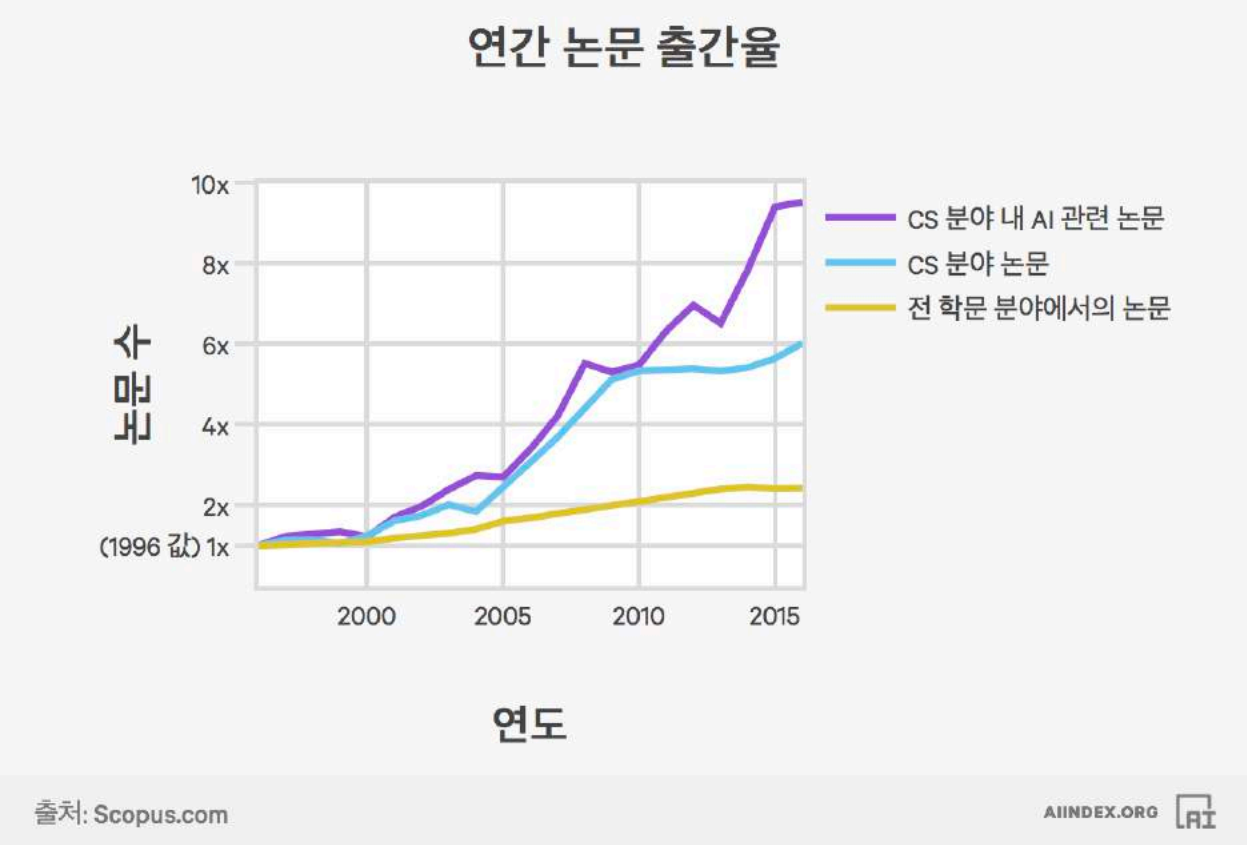
부록 A1 에서 추가 정보 보기

SCOPUS 논문 인용 분석 데이터베이스에서 '인공지능(Artificial Intelligence)'을 키워드로 태그되어 게재된 컴퓨터 과학 논문 수.



9x 매년 출간되는 인공지능 관련 논문 수
*1996*년 이후 *9*배 이상 증가.

1996년 대비 연간 논문 출간율 비교. 이하 그래프는 전 학문 분야에서의 논문 출간 증가율, 컴퓨터 과학 분야 논문 출간 증가율, 컴퓨터 과학 분야 내 인공지능 관련 논문 출간 증가율이 모두 우상향 추세임을 보여준다.



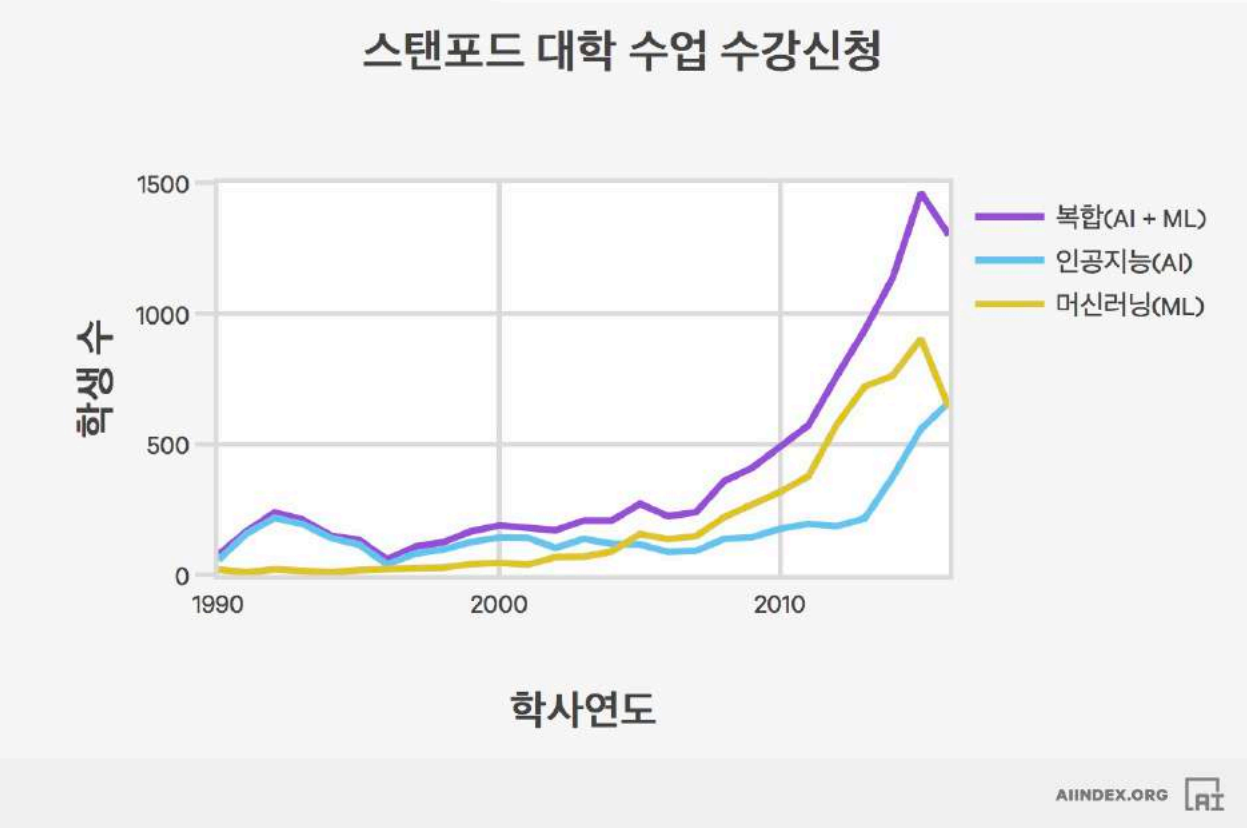
본 데이터는 인공지능 관련 논문의 증가가 컴퓨터 과학 분야 전반에 대한 관심이 증가한 것 이상의 원인에 기인하고 있음을 보여준다. 구체적으로 1996년 이후 매년 컴퓨터 과학 분야에서 출간된 논문은 6배 증가한 반면 인공지능 관련 논문은 동기간 9배 이상 증가하였다.

수강 신청

부록 A2 에서 추가 정보 보기

스탠포드 대학교에서 인공지능 및 기계학습 개론 수업에 등록한 학생 수.

기계학습(Machine Learning)은 인공지능(Articifial Intelligence)의 세부 분야이며, 본 보고서에서 기계학습 강의를 다루는 이유는 기계학습 강의 수강생의 급증과 최근 인공지능 기술 발전에서 기계학습이 핵심적인 역할을 하기 때문이다.

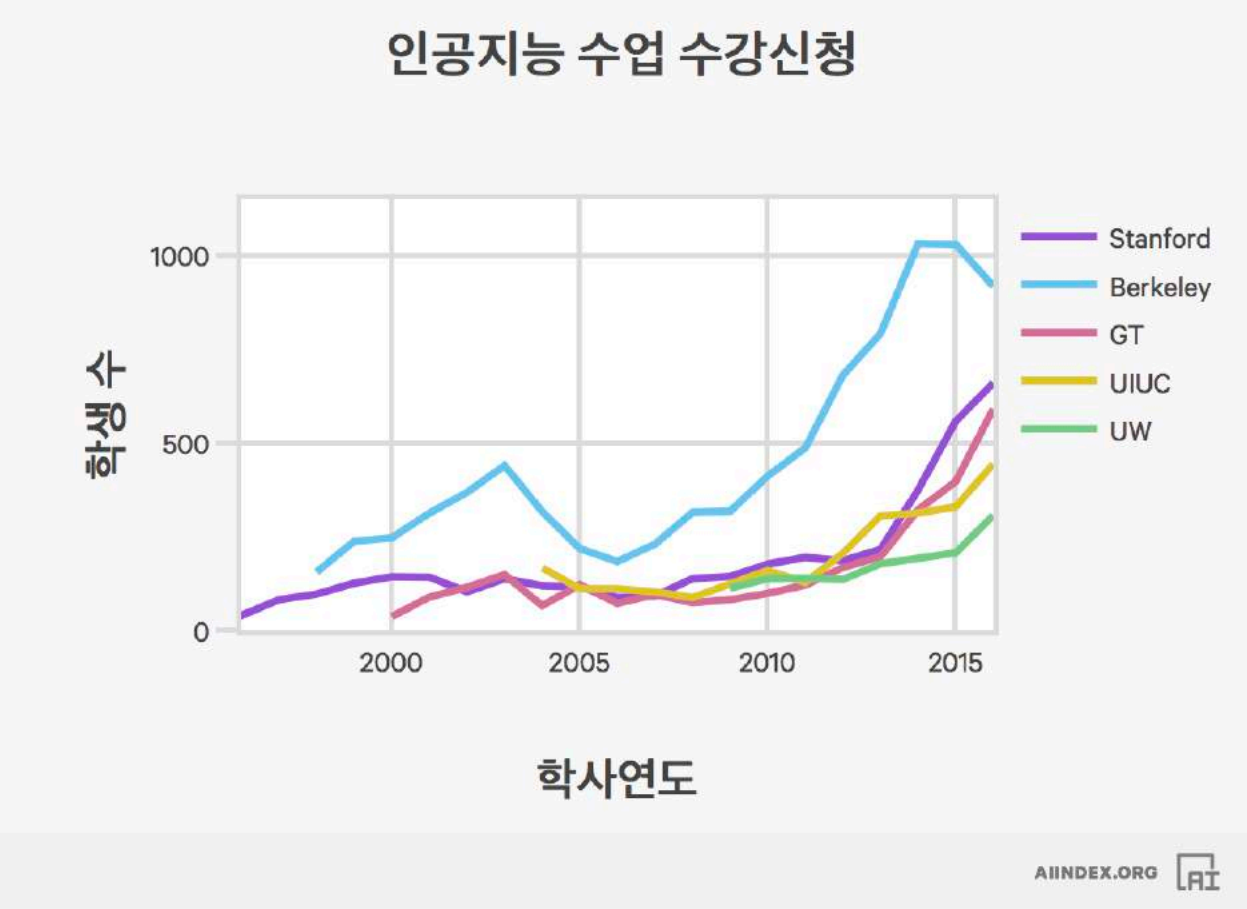


11x

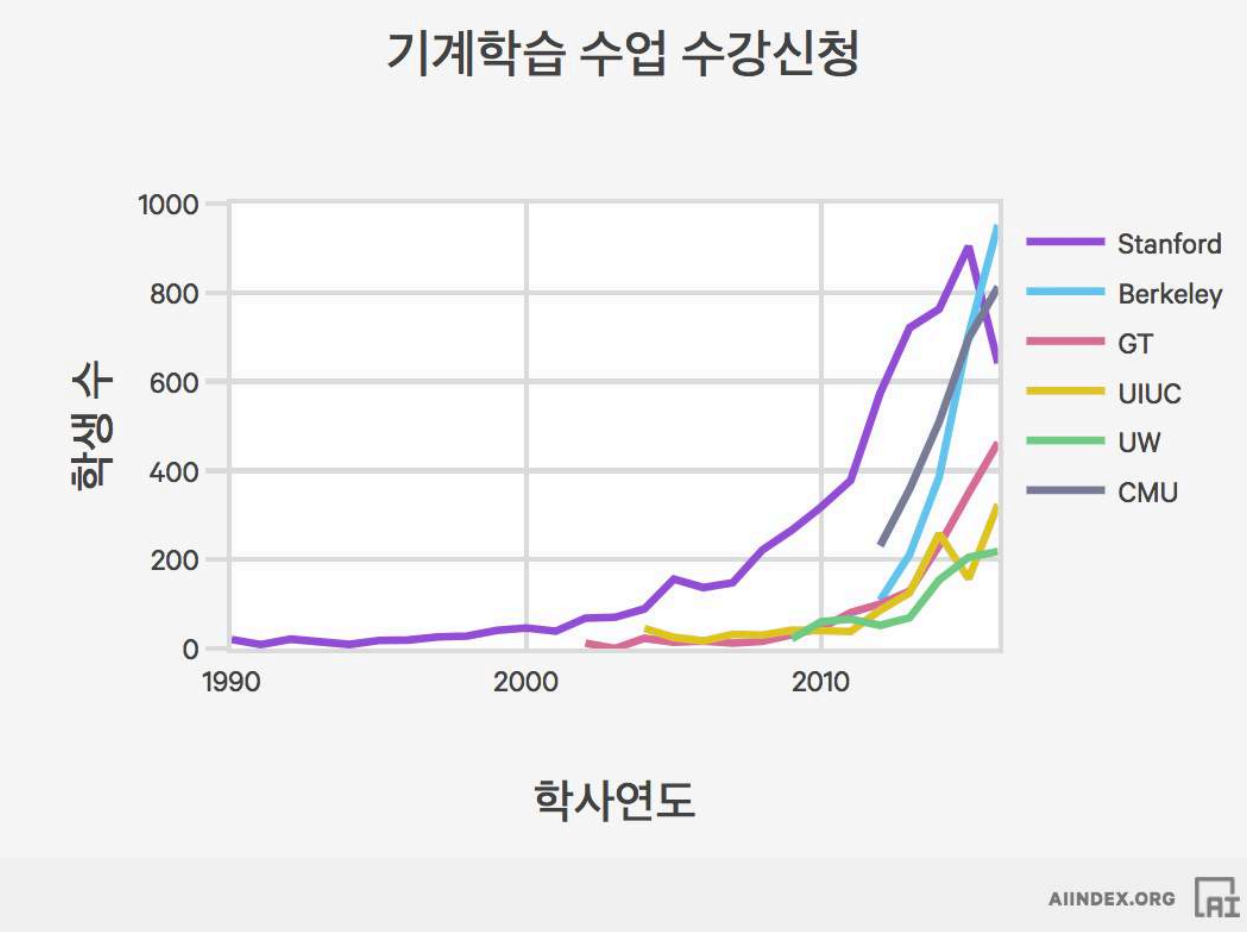
스탠포드 대학교에서 인공지능 개론 수업 수강신청 학생 수 *1996*년 이후 *11*배 증가.

참고 : 2016 학년도에 스탠포드 대학교 기계학습 수업 수강신청 학생 수 하락은 학생들의 흥미 하락으로 인한 것이 아닌 해당 학년도의 행정적 원인에 의한 것이며 자세한 내용은 부록 참고 요망.

본 보고서에서 스탠포드 대학교에 초점을 맞춘 이유는 타 대학교의 경우 관련 데이터가 제한적이기 때문이다. 그러나 타 대학교 학생들의 과거 수강신청 경향이 스탠포드 대학교 학생들이 보인 경향과 유사하다고 추정할 수 있다.



참고 : 많은 대학에서 90년대 이전부터 인공지능 관련 강의를 제공했다. 위 그래프는 가용 데이터가 존재한 연도를 나타낸다.



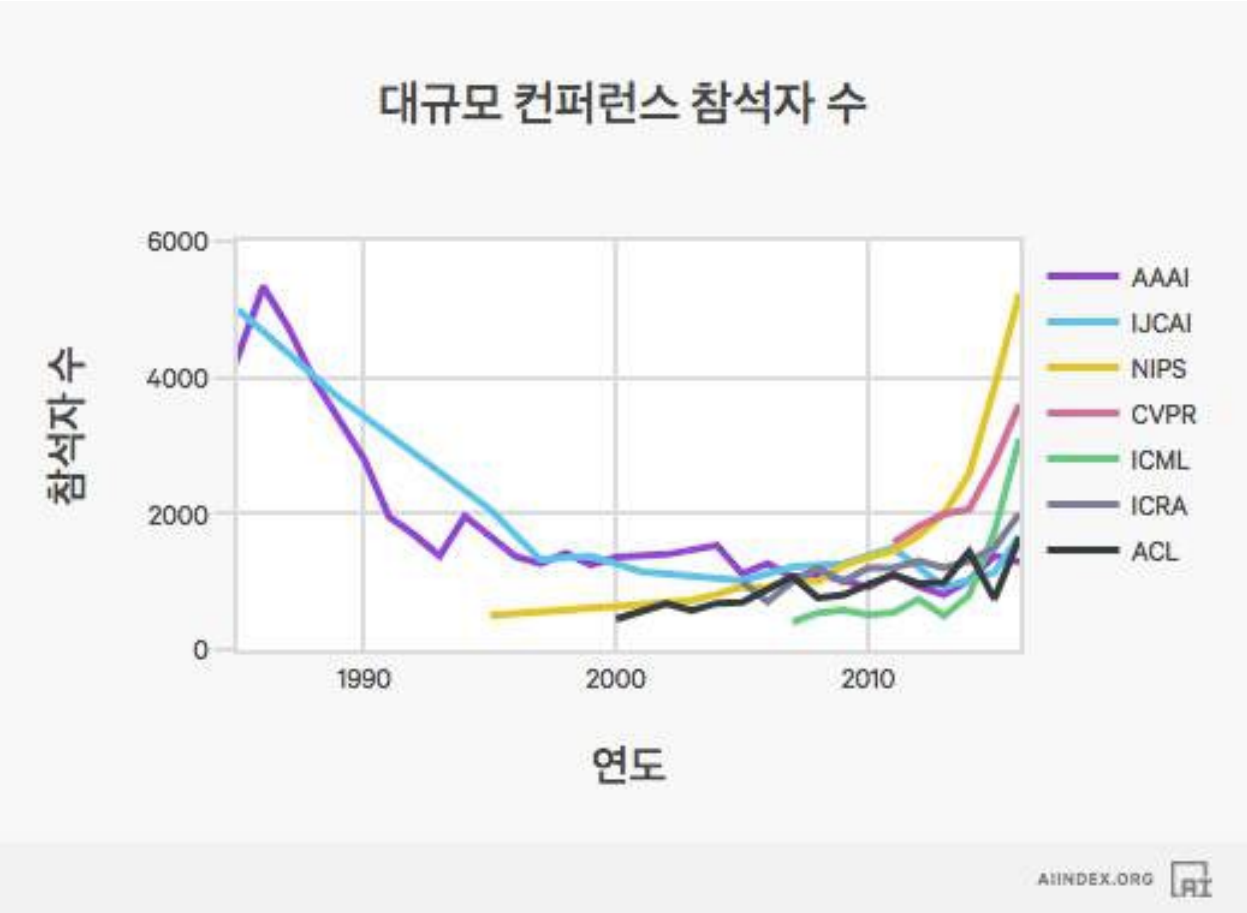
참고 : 많은 대학에서 90년대 이전부터 기계학습 관련 강의를 제공했다. 위 그래프는 가용 데이터가 존재한 연도를 나타낸다.

위 그래프는 고등 교육에서 보인 경향의 일부를 보여줄 뿐이며 학문 기관 전반을 대변하지는 않을 수 있다.

컨퍼런스 참석자

부록 A3 에서 추가 정보 보기

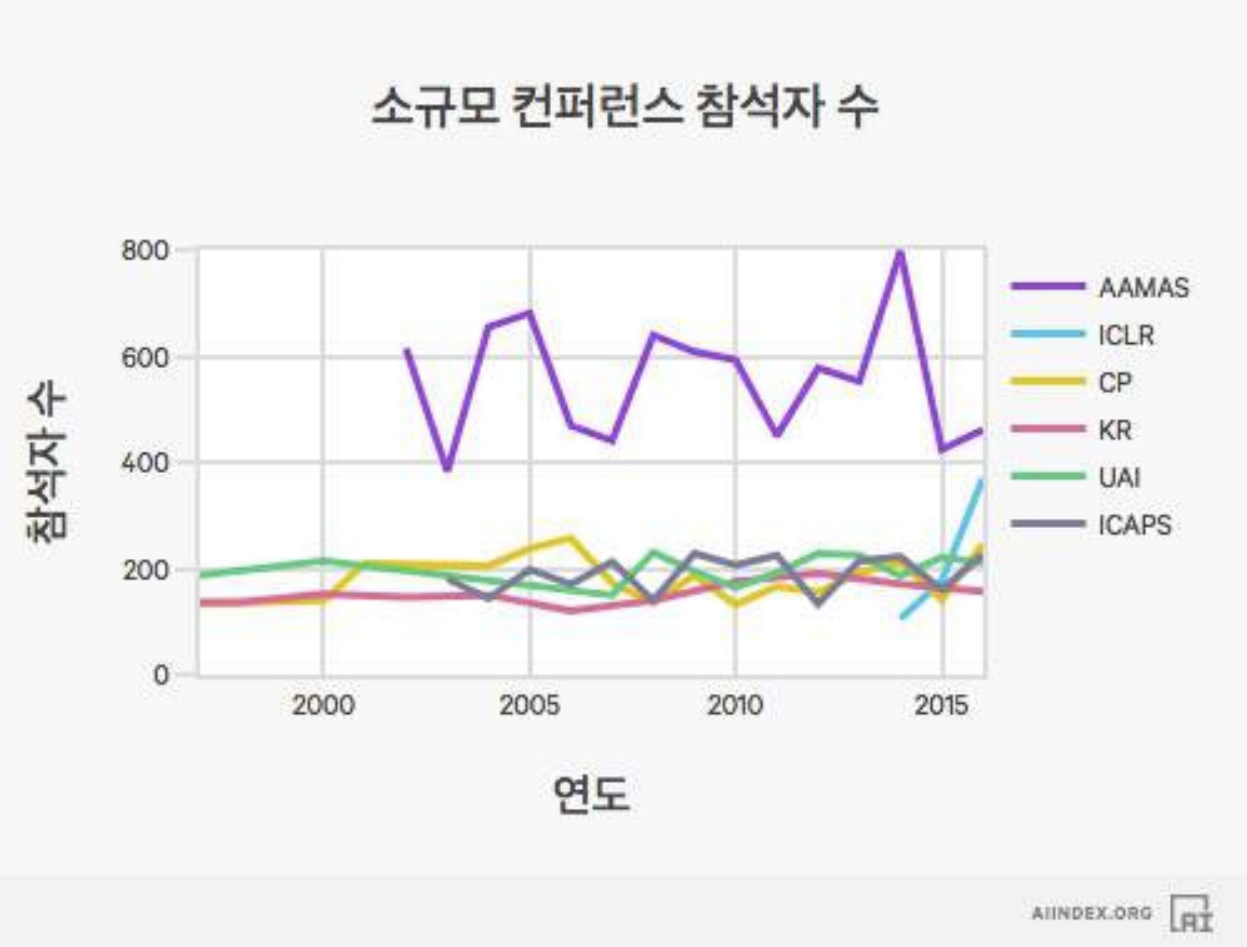
주요 인공지능 컨퍼런스 참석자 수. 본 데이터는 대규모 컨퍼런스(2016년 기준 참석자 수 1000명 초과)와 소규모 컨퍼런스(2016년 기준 참석자 수 1000명 미만)로 나누어져 있다.



연구주제 변화

컨퍼런스 참석자 수는 연구 주제가 상징 추론에서 기계학습과 딥러닝으로 옮겨갔음을 보여준다.

참고 : 대부분의 컨퍼런스는 1980년대부터 존재했다. 위 데이터는 참석자 데이터가 기록된 연도의 수치를 나타낸다.



지속적인 발전

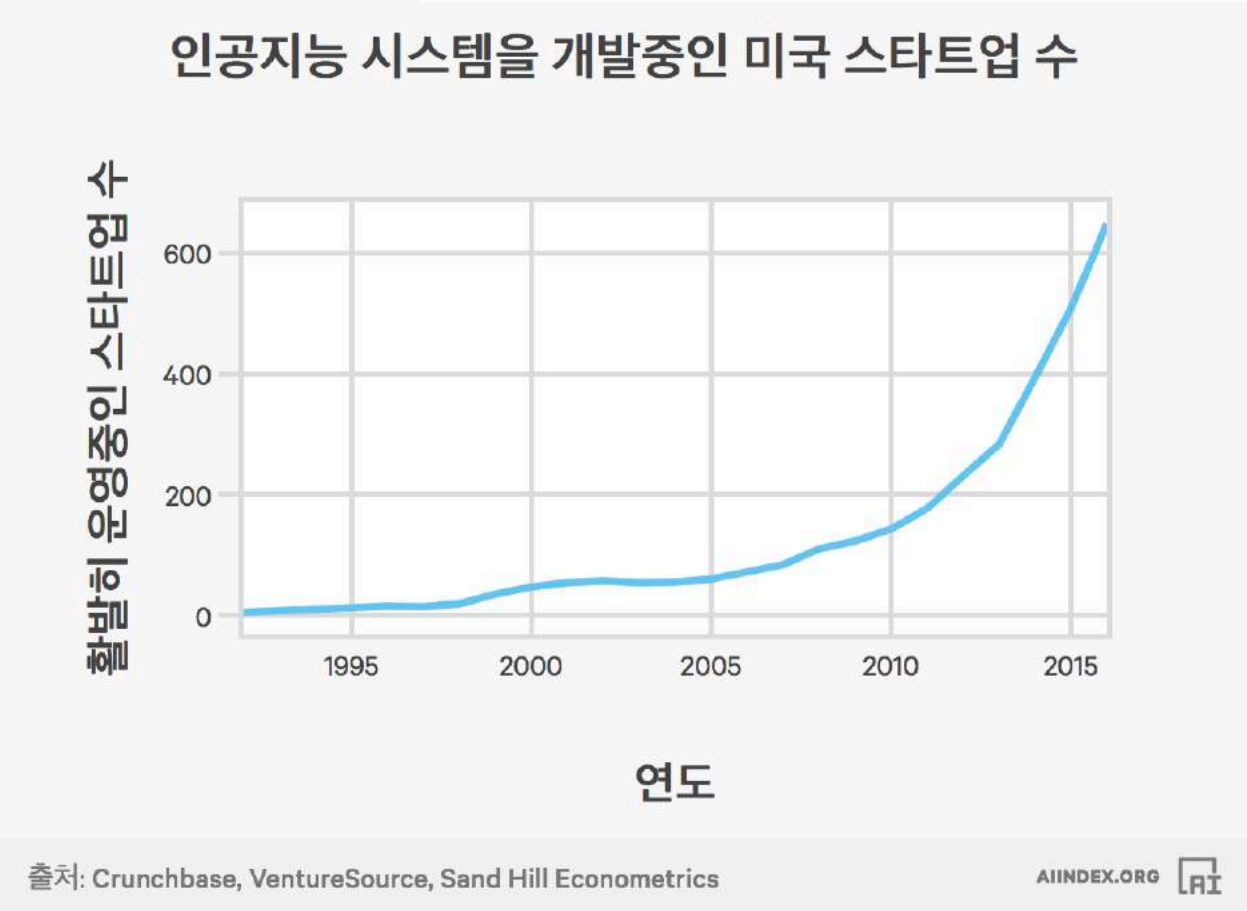
주요 연구주제는 변화하고 있지만 인공지능에서 상징 추론 방식을 지속적으로 발전시키는 소규모 연구 커뮤니티 존재.

산업계

인공지능 관련 스타트업

부록 A4 에서 추가 정보 보기

벤처캐피털 투자 기반으로 인공지능 시스템을 개발 중이며 활발히 운영 중인 비상장 미국 스타트업의 수.

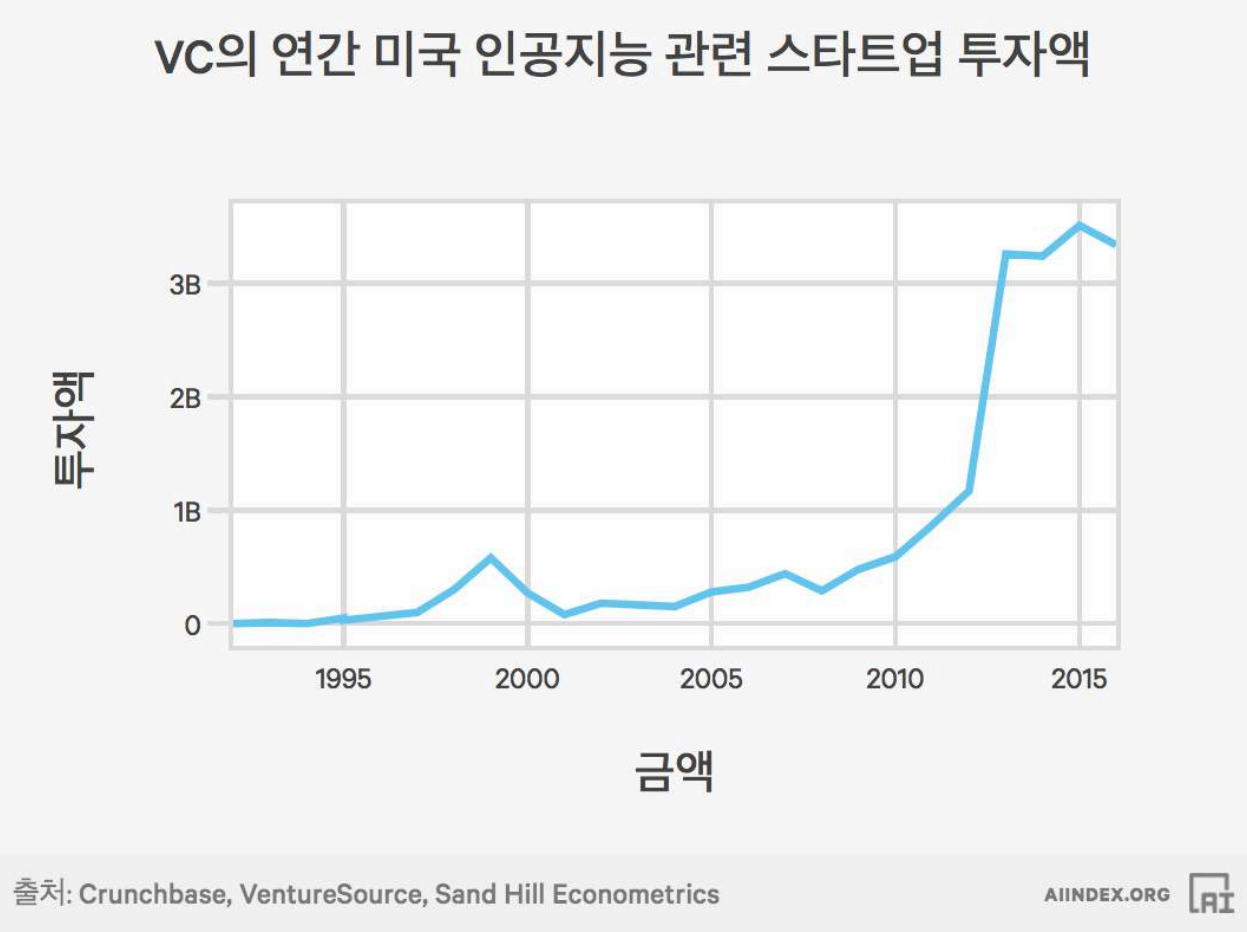


14x 인공지능 시스템을 개발 중인 활발히 운영 중인 미국 스타트업의 수 *2000*년 이후 *14*배 증가.

인공지능 관련 스타트업 투자

부록 A5 에서 추가 정보 보기

모든 투자 단계에서 벤처캐피털(VC)이 연간 미국 인공지능 관련 스타트업에 투자하는 금액.



6x 미국 인공지능 스타트업에 대한 연간 벤처 투자 금액 *2000*년 이후 *6*배 증가.

일자리

부록 A6 에서 추가 정보 보기

인공지능 관련 일자리 증가율 관련 데이터는 온라인 구인 플랫폼 인디드(Indeed)와 몬스터(Monster)를 통해 얻었다. 인공지능 관련 일자리는 직무 설명에서 제목과 키워드를 통해 판별되었다.

Indeed 플랫폼에서 인공지능 기술을 필요로 하는 미국 일자리 비중이 커지고 있음. 이하 그래프는 2013년 1월 미국 기준 Indeed 플랫폼에서 인공지능 기술을 필요로 하는 일자리의 증가세를 배수로 나타내었다.

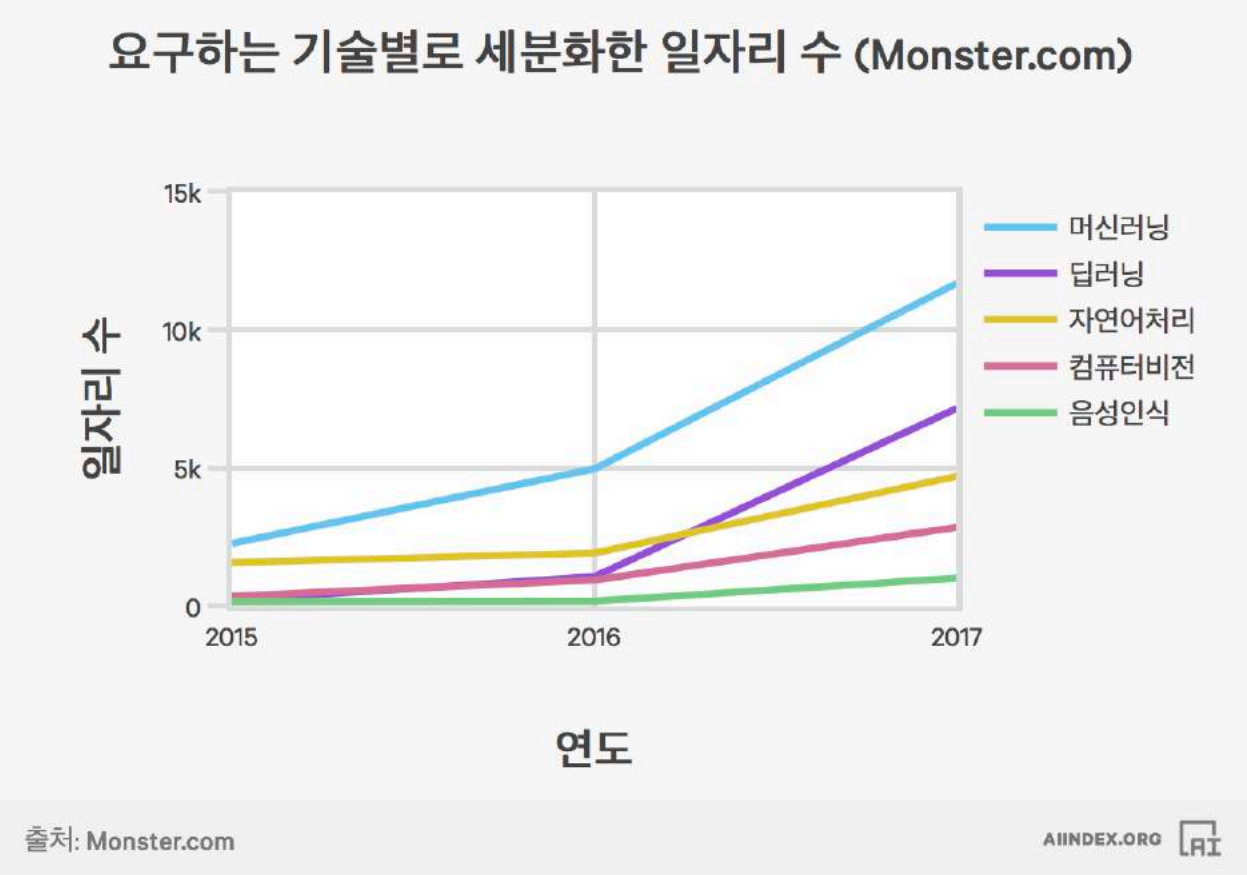
4.5x 미국에서 인공지능 기술을 필요로 하는 직업

2013년 이후 4.5배 증가.

Indeed 플랫폼에서 국가별로 인공지능 기술을 필요로 하는 직업의 증가세.

참고 : 캐나다와 영국에서 인공지능 관련 일자리가 급속하게 증가하였음에도, Indeed에 따르면 각각 미국 인공지능 관련 고용 시장의 절대 크기 대비 5%와 27%에 불과하다.

해당 연도에 Monster 플랫폼에 게시된 총 인공지능 일자리 수를 요구하는 기술별로 세분화.

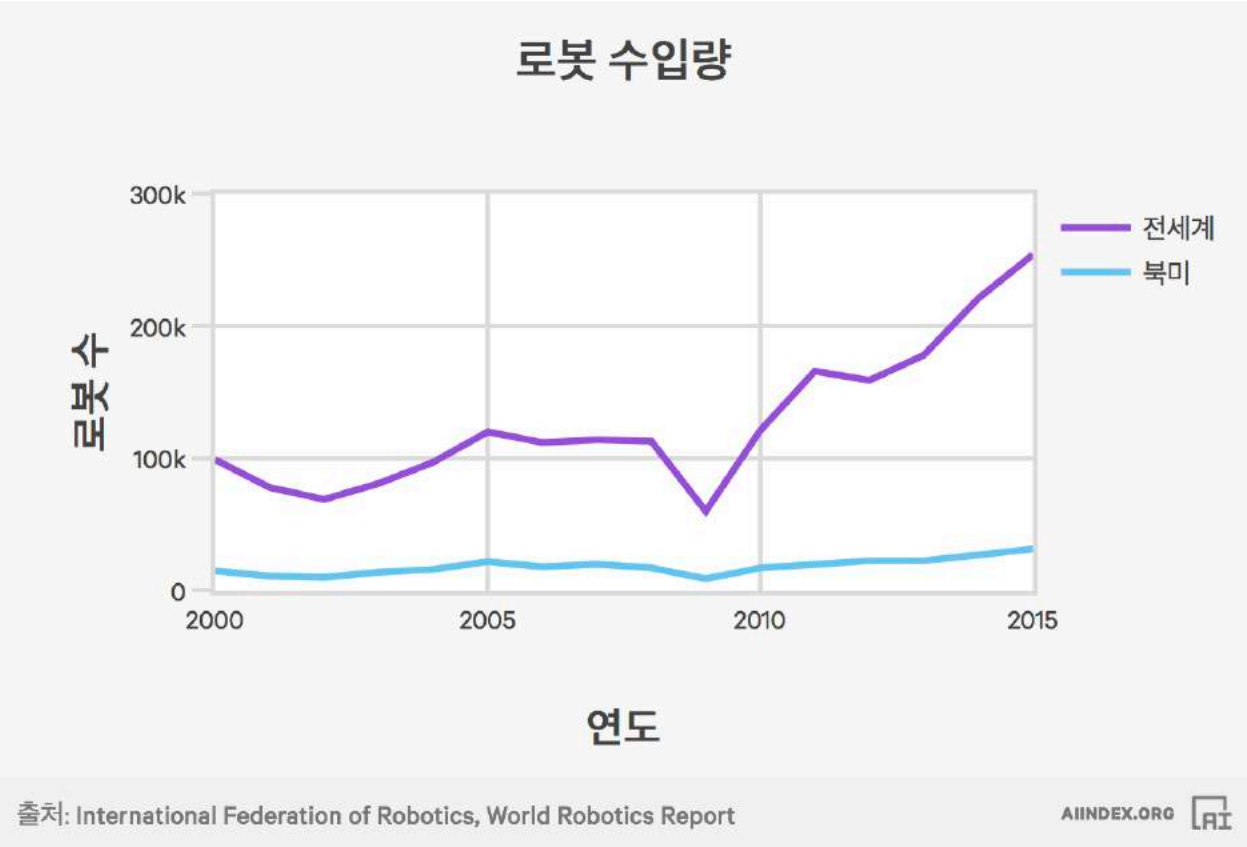


참고 : 단일 인공지능 관련 일자리는 두개 이상의 범주에 포함될 수 있다. 예를 들어 하나의 일자리가 자연어처리와 컴퓨터비전 기술을 모두 요구할 수도 있다.

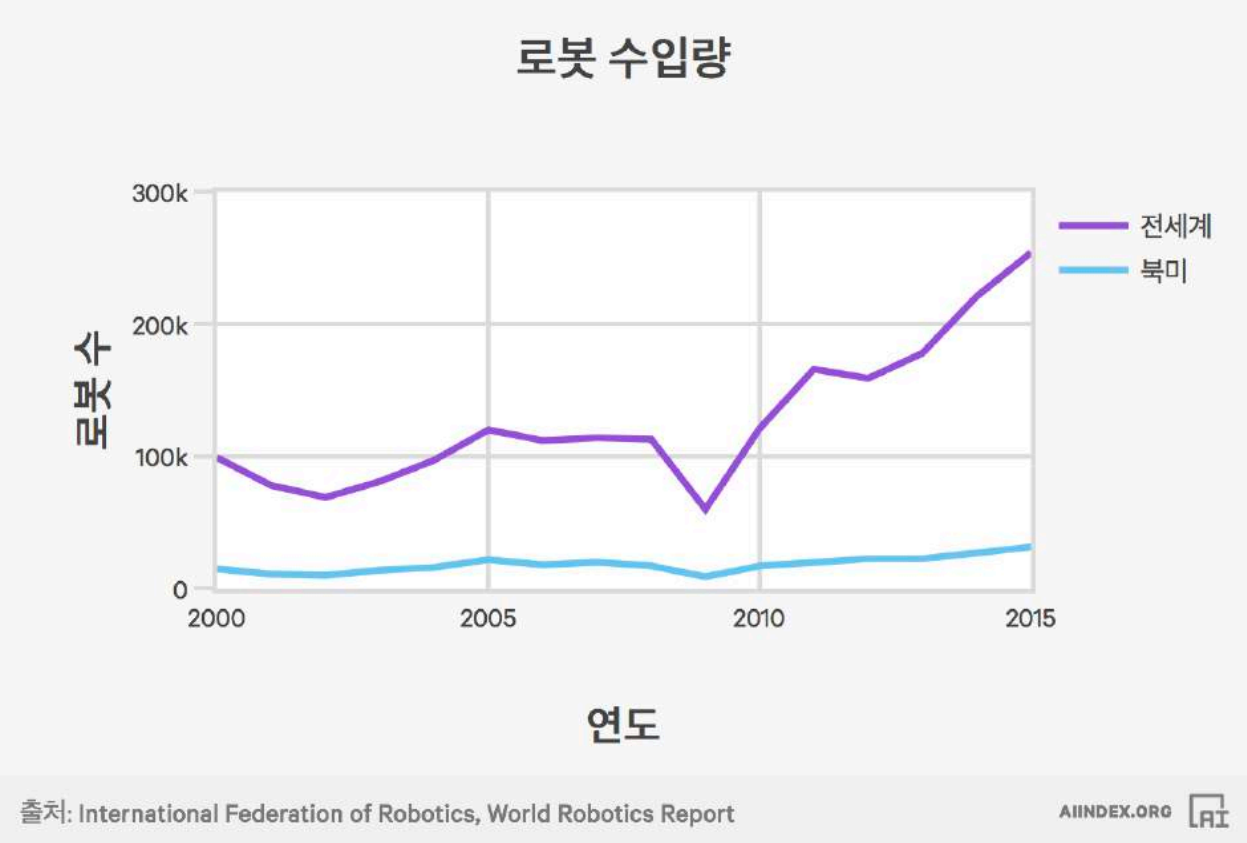
로봇 수입량

부록 A7 에서 추가 정보 보기

북미 및 전 세계적 산업용 로봇 수입량.



북미 및 전 세계적 산업용 로봇 유닛 수입 증가율.

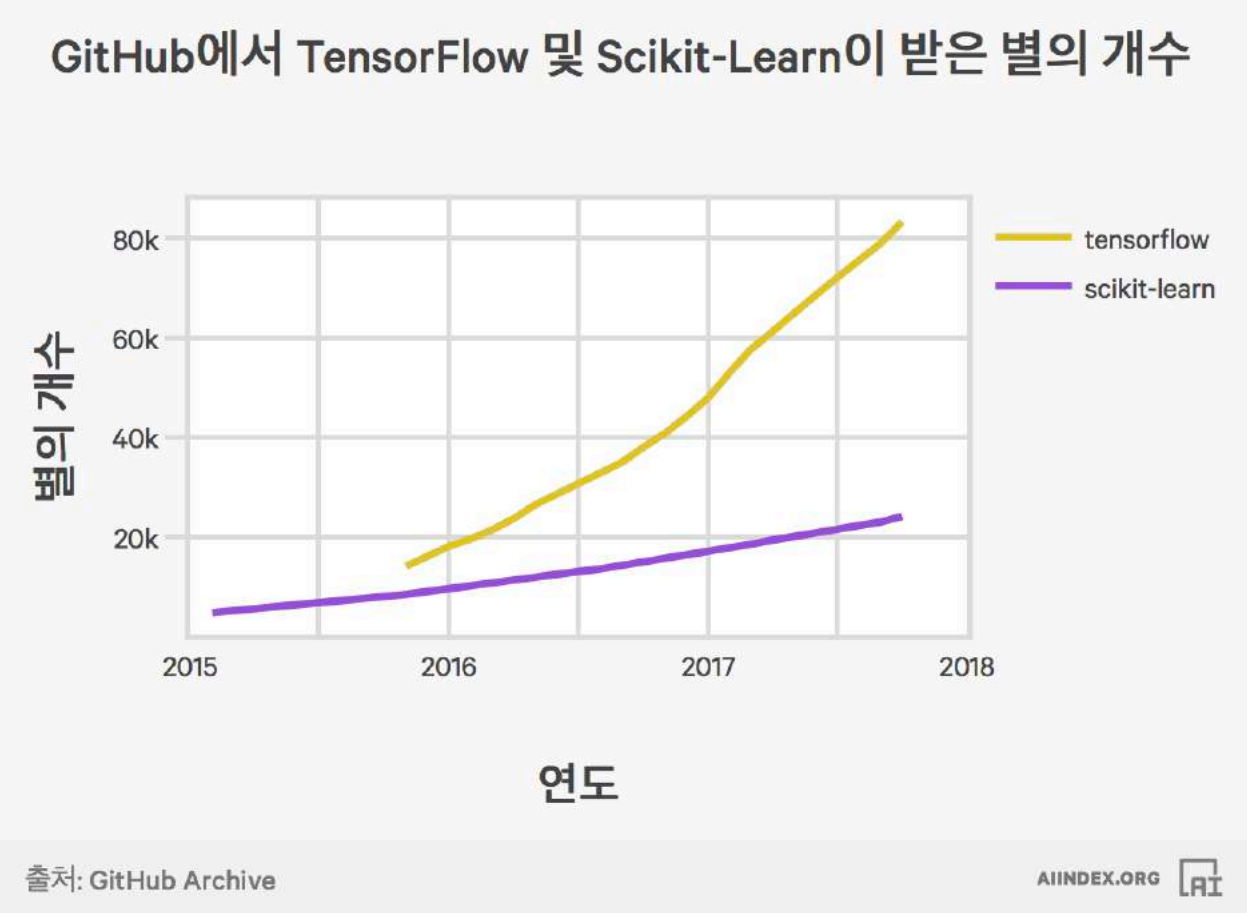


오픈소스 소프트웨어

깃허브(GitHub) 프로젝트 통계

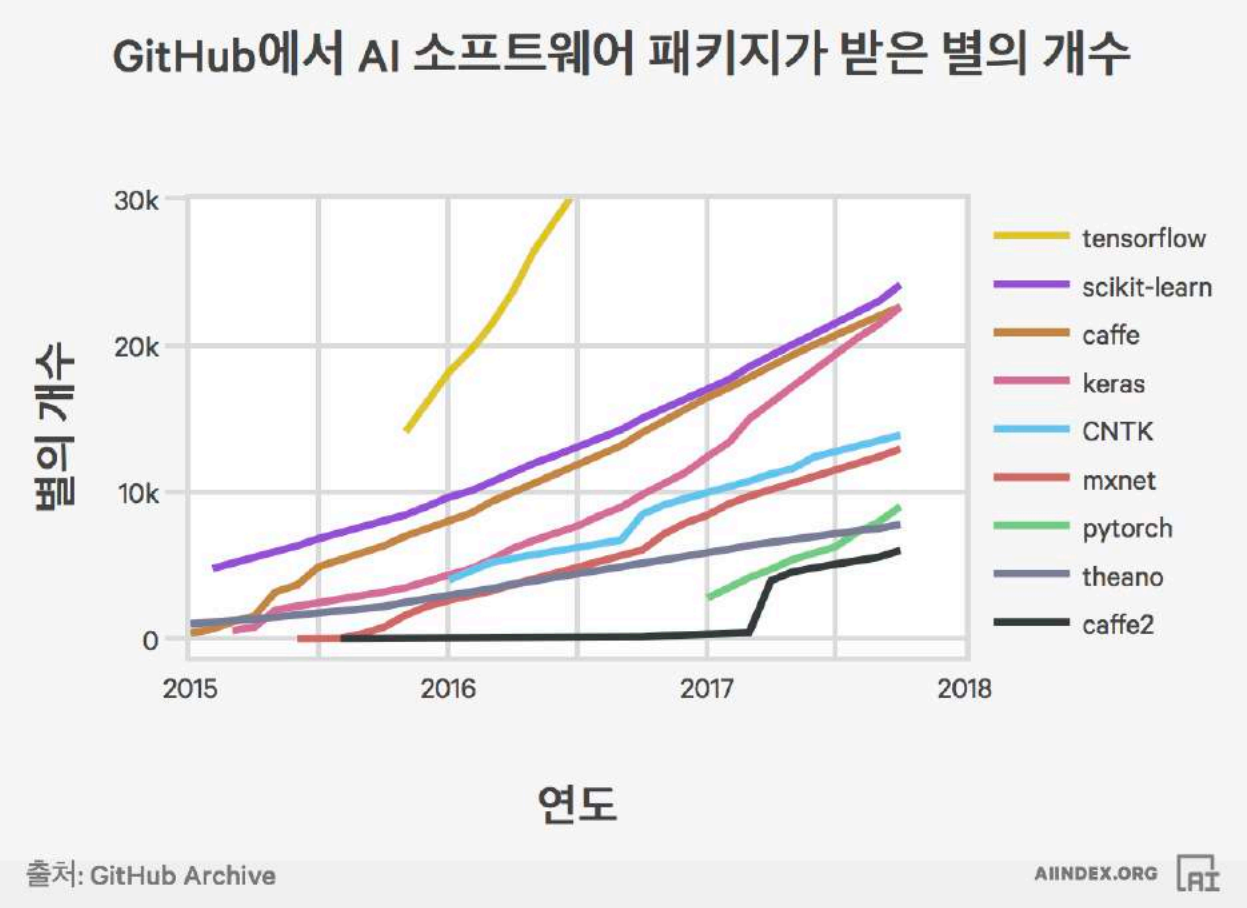
부록 A8 에서 추가 정보 보기

깃허브(GitHub) 사용자들이 텐서플로우(TensorFlow) 및 사이킷런(Scikit-Learn) 소프트웨어 패키지에 매긴 별의 수. TensorFlow 와 Scikit-Learn은 딥러닝 및 기계학습과 관련하여 인기 있는 소프트웨어 패키지이다.



소프트웨어 개발자들은 GitHub에서 흥미롭거나, 감사를 표하고 싶거나, 신속하게 탐색하고 싶은 프로젝트에 '별(Star)'을 줄 수 있다. 개발자들이 프로젝트에 매긴 별의 개수를 통해 각 프로젝트에 대한 개발자들의 관심도와 해당 소프트웨어의 사용 정도를 가늠해 볼 수 있다.

GitHub에서 인공지능 및 기계학습 소프트웨어 패키지가 받은 별의 개수.



참고 : 깃허브(GitHub) 저장소의 포크(Fork)는 별의 개수와 거의 동일한 경향을 보인다(각 저장소마다 Fork와 별의 절대 개수는 다르다). Fork 데이터 수집에 대한 정보는 부록 참조 요망.

대중적 관심

언론 보도 동향

부록 A9 에서 추가 정보 보기

'인공지능(Artificial Intelligence)'이라는 단어가 포함된 언론 기사의 비율을 긍정적 또는 부정적 기사로 분류.

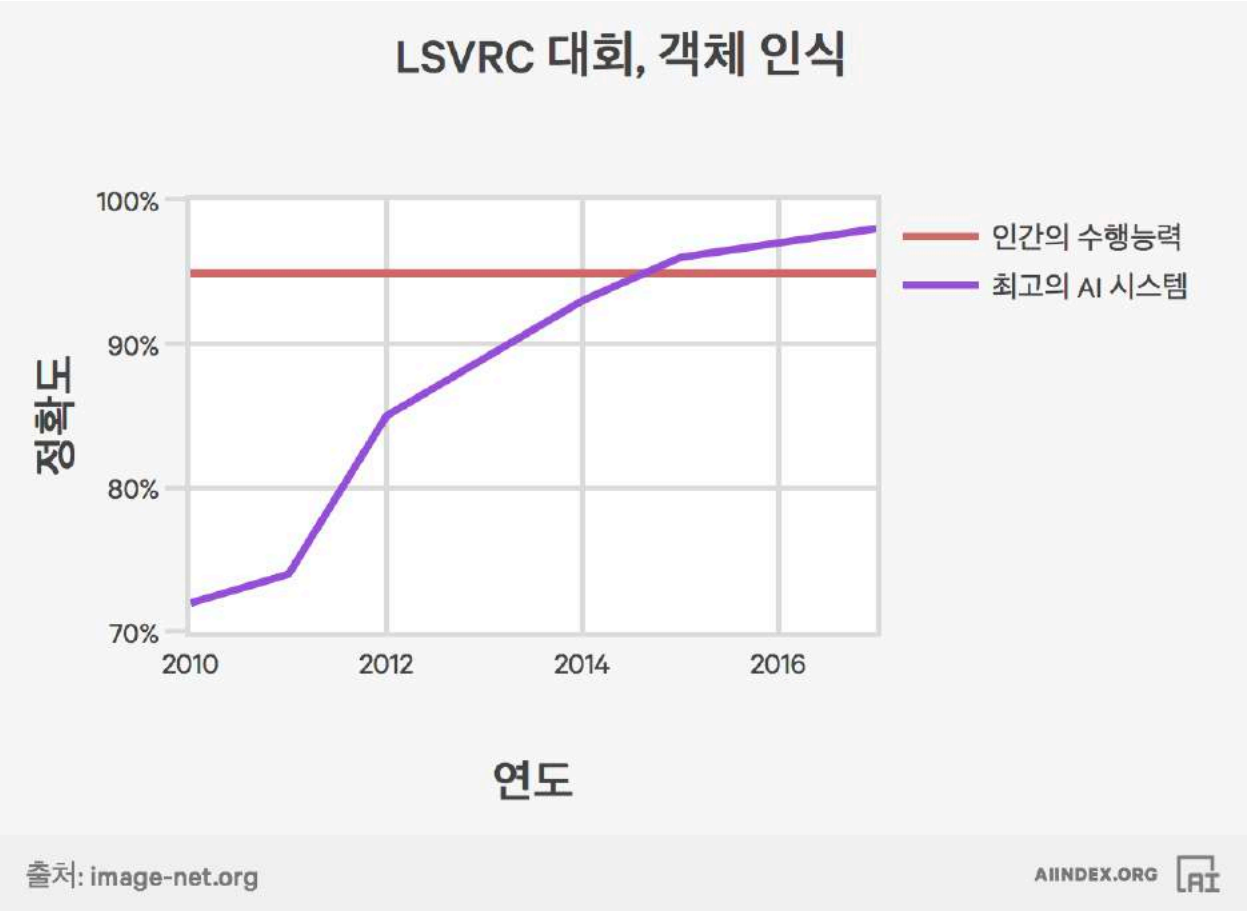
기술적 성능

시각(Vision)

객체 인식(Object Detection)

부록 A10 에서 추가 정보 보기

대용량 시각 인식 대회(Large Scale Visual Recognition Challenge, LSVRC) 대회에서 객체 인식 작업 부문 인공지능 시스템별 성능.

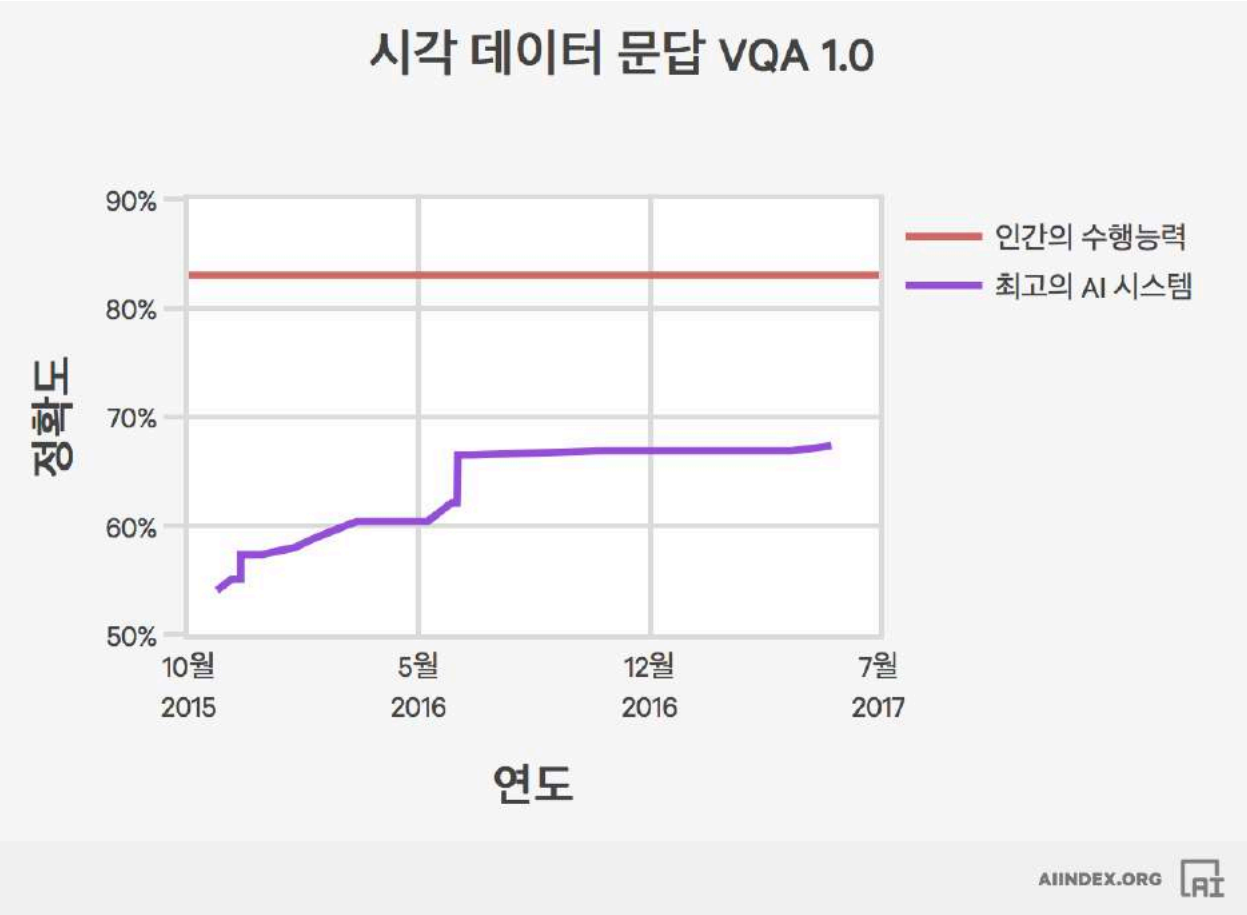


2.5% *2010*년 이후 이미지 레이블링에 대한 오류율 *28.5%*에서 *2.5%* 미만으로 감소.

시각 데이터 문답(Visual Question Answering)

부록 A11 에서 추가 정보 보기

이미지에 대한 질문에 열린 답변(Open-ended answers)을 내놓는 작업에서 인공지능 시스템의 성능.



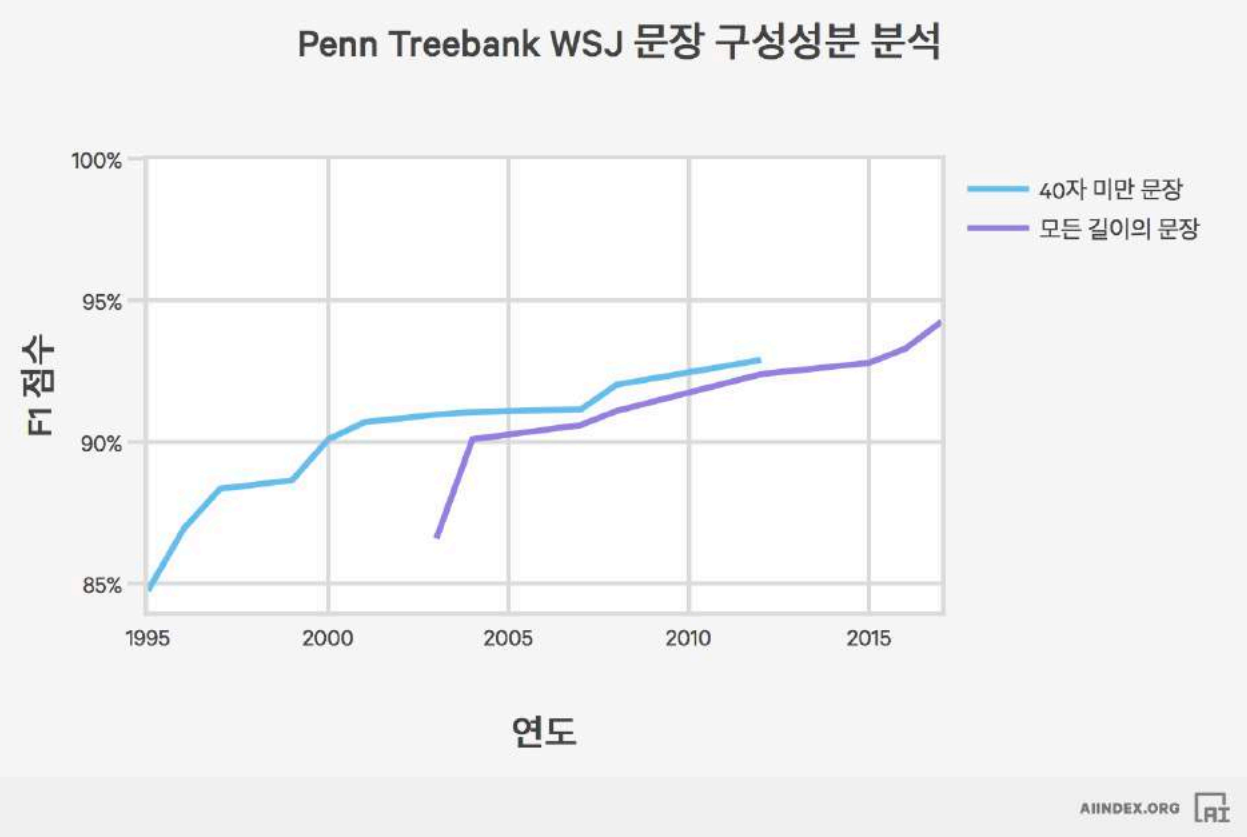
참고 : *VQA 2.0* 데이터 셋은 이미 *VQA 1.0* 데이터 셋을 능가하였으며, *VAQ 1.0* 데이터 셋에 사람들이 얼마나 많은 관심을 기울일지는 명확하지 않음.

자연어 이해(Natural Language Understanding)

파싱(Parsing)

부록 A12 에서 추가 정보 보기

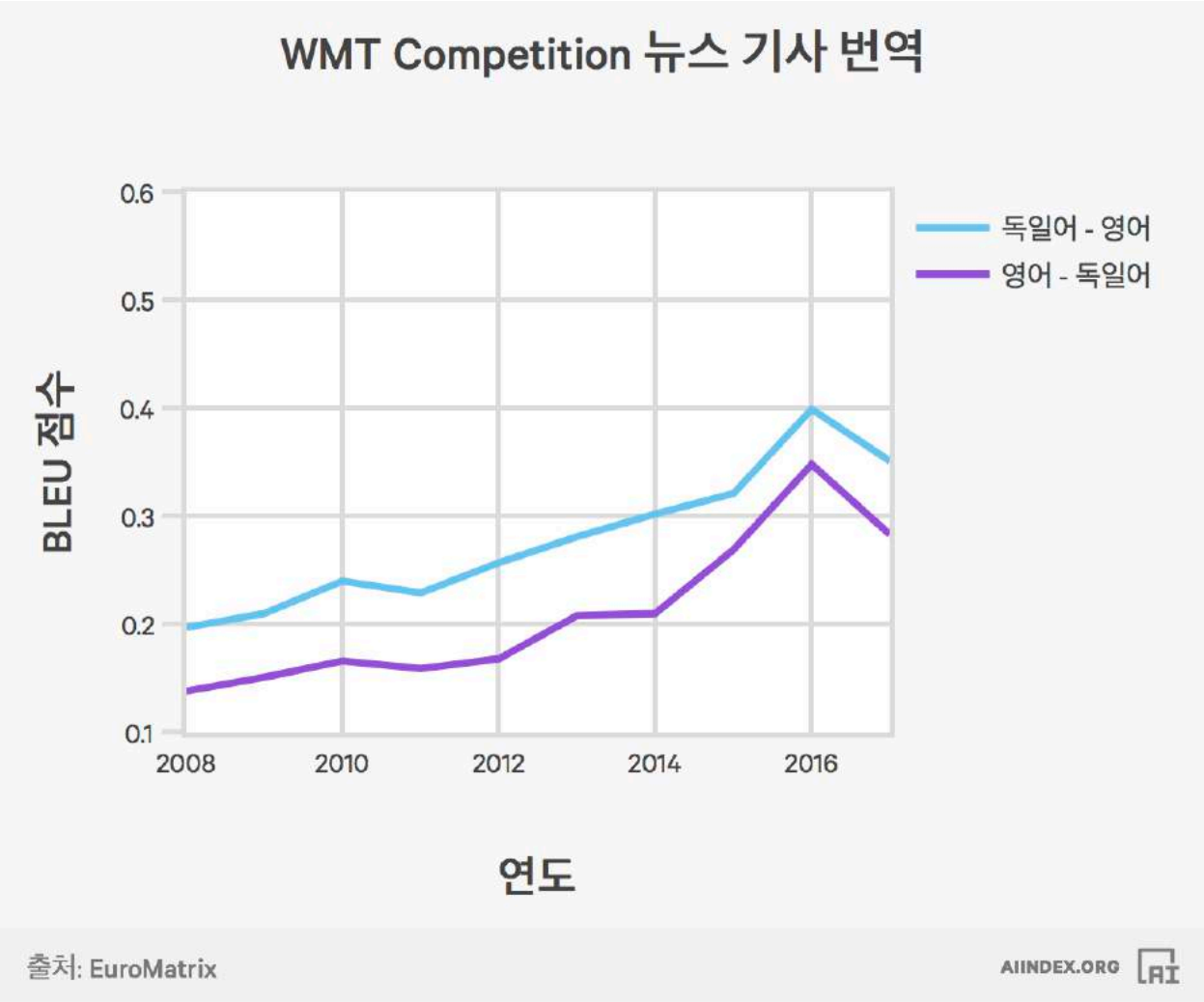
인공지능 시스템의 문장 구문 구조(syntactic structure) 분석 성능.



기계 번역(Machine Translation)

부록 A13 에서 추가 정보 보기

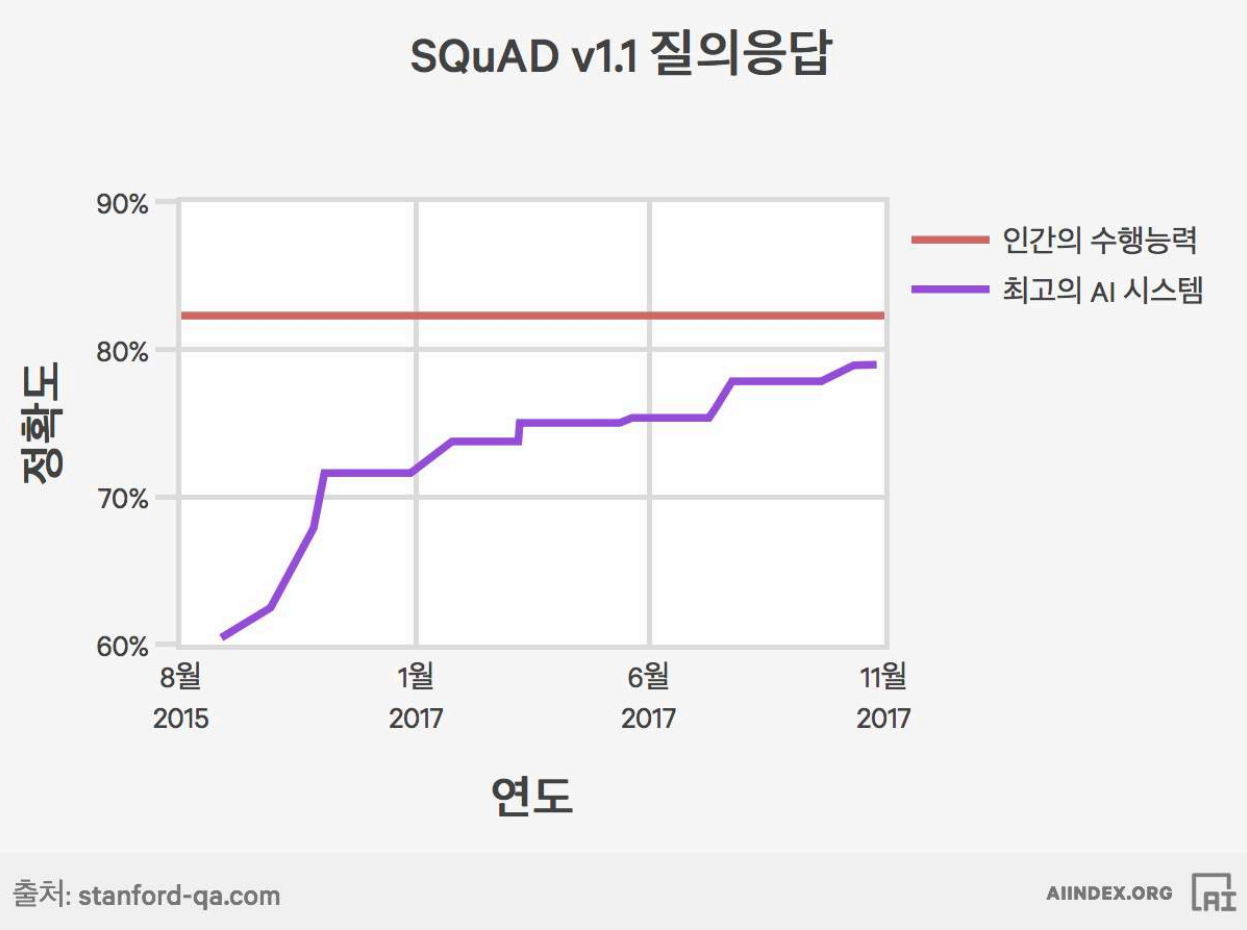
인공지능 시스템의 영어-독일어, 독일어-영어 뉴스 기사 번역 성능.



질의응답(Question Answering)

부록 A14 에서 추가 정보 보기

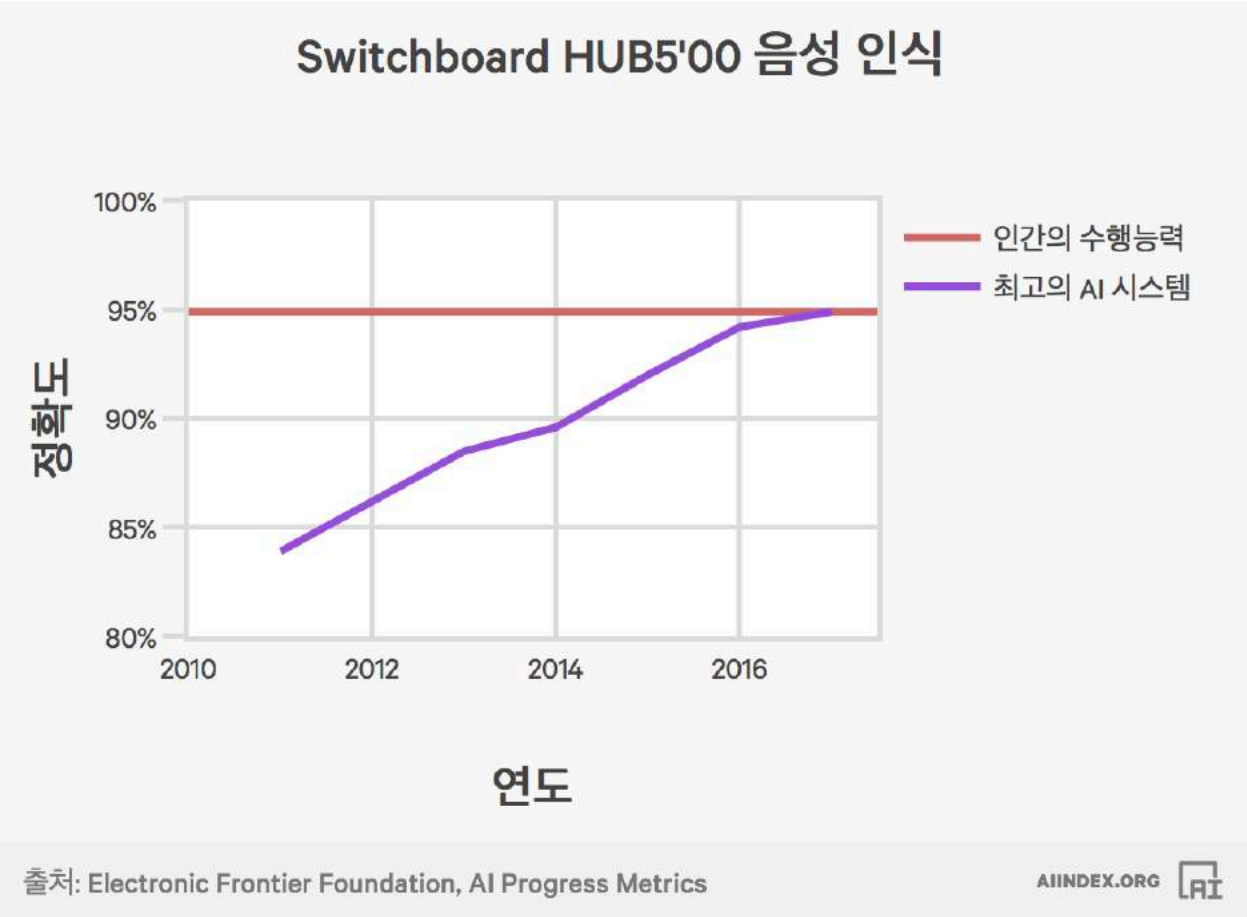
인공지능 시스템의 질문에 대해 문서 내에서 답변을 찾는 성능.



음성 인식(Speech Recognition)

부록 A15 에서 추가 정보 보기

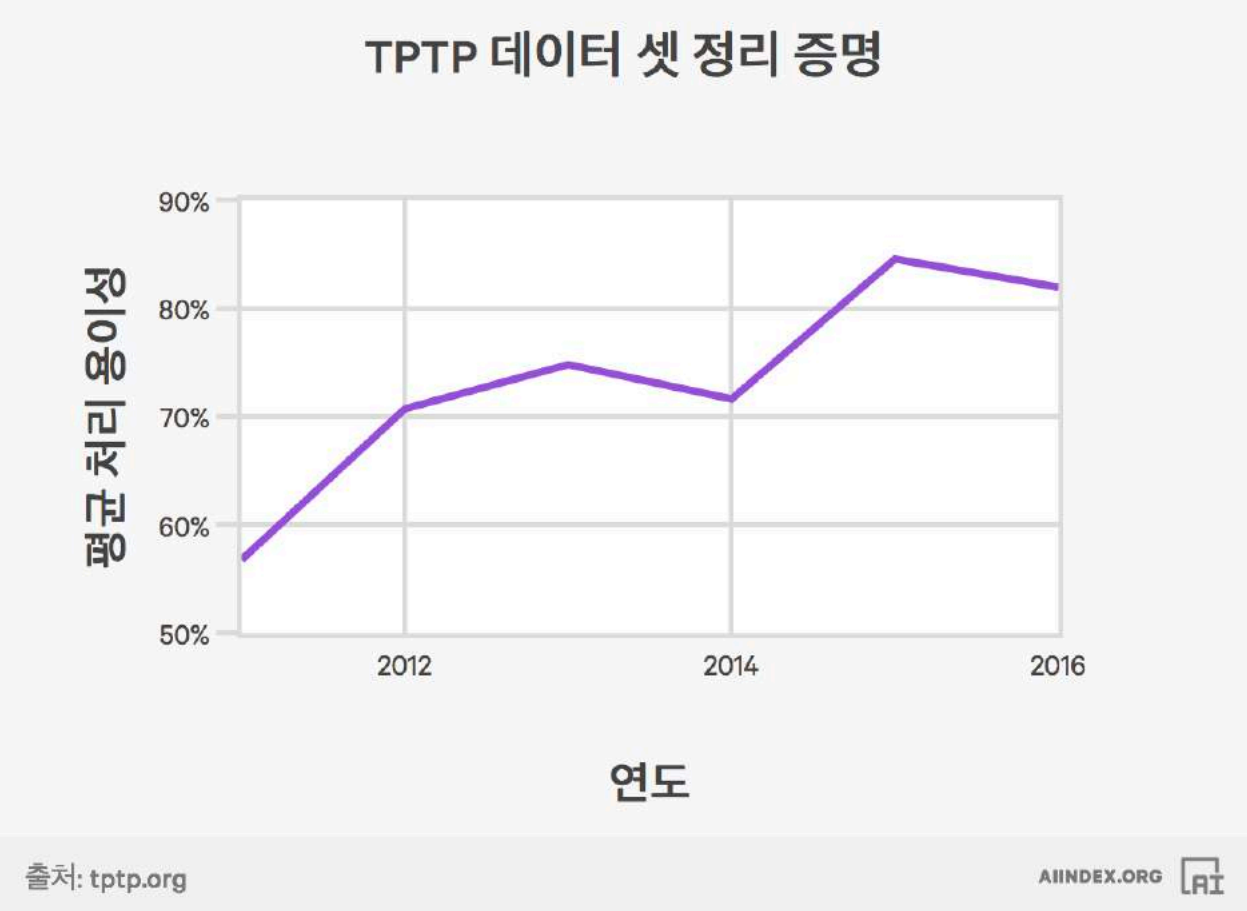
인공지능 시스템의 전화 통화 소리에서 음성을 인식하는 성능.



정리 증명(Theorem Proving)

부록 A16 에서 추가 정보 보기

자동 정리 증명기(Automatic Theorem Provers)를 위한 대규모 정리 증명 문제(theorem proving problems)의 평균 처리 용이성(Tractability). '처리 용이성' 이란 문제를 풀 수 있는 최신 자동 정리 증명기의 비율을 의미한다. '처리 용이성' 지표에 관한 구체적인 정보는 부록 참고 요망.

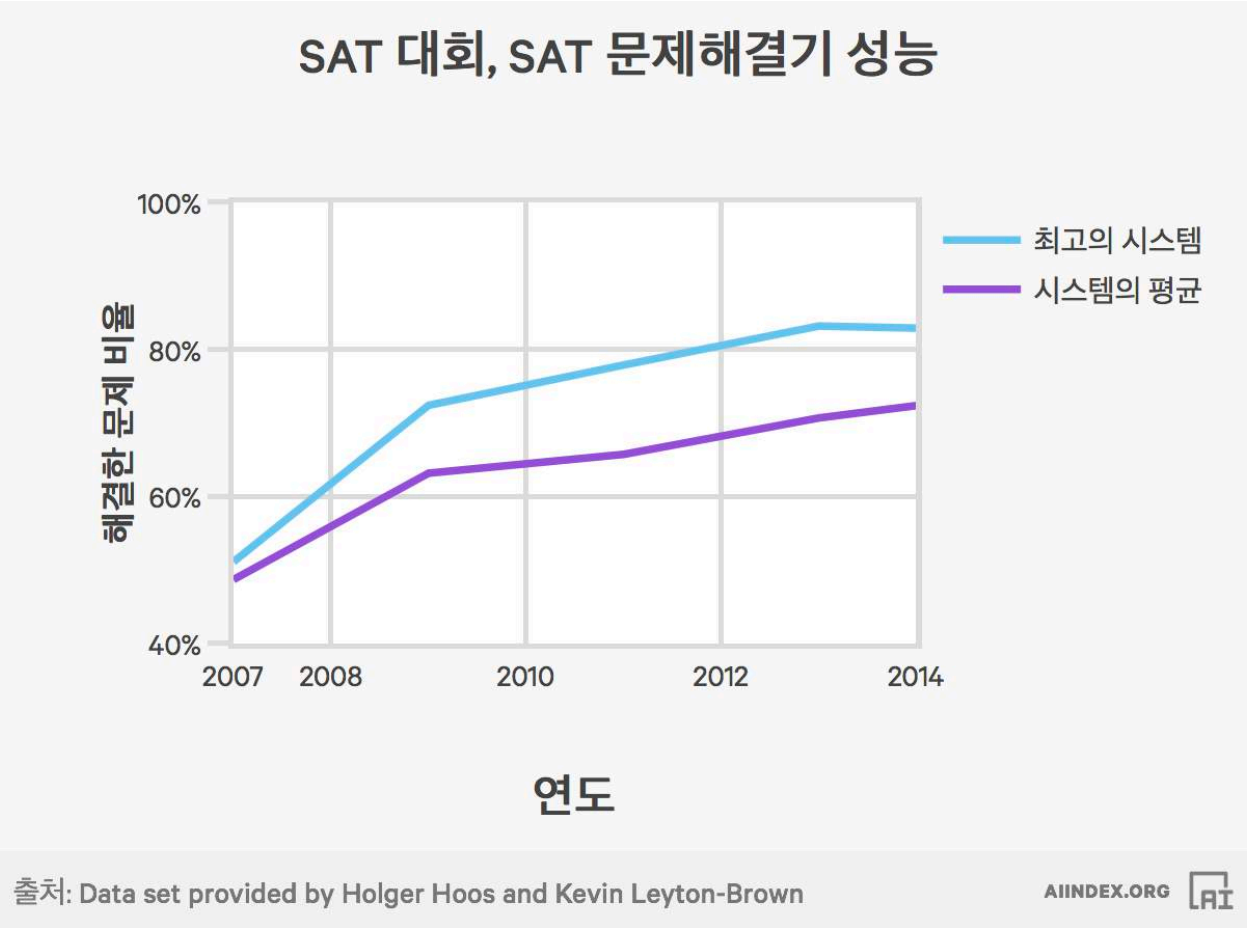


참고 : 최신 문제해결기(*solver*)가 새로운 문제를 해결하는데 좋은 성능을 보이지만, 반면에 다른 문제해결기(*solver*)가 잘 해결하는 문제에 좋지 않은 성능을 보인다면 평균적인 처리 용이성은 감소할 수 있다.

충족 가능성 문제 해결(SAT Solving)

부록 A17 에서 추가 정보 보기

산업적 응용이 가능한 문제 해결에 탁월한 SAT 문제해결기(Solver)가 해결한 문제 비율.



출처: Data set provided by Holger Hoos and Kevin Leyton-Brown

AIINDEX.ORG

파생된 지표(DERIVATIVE MEASURES)

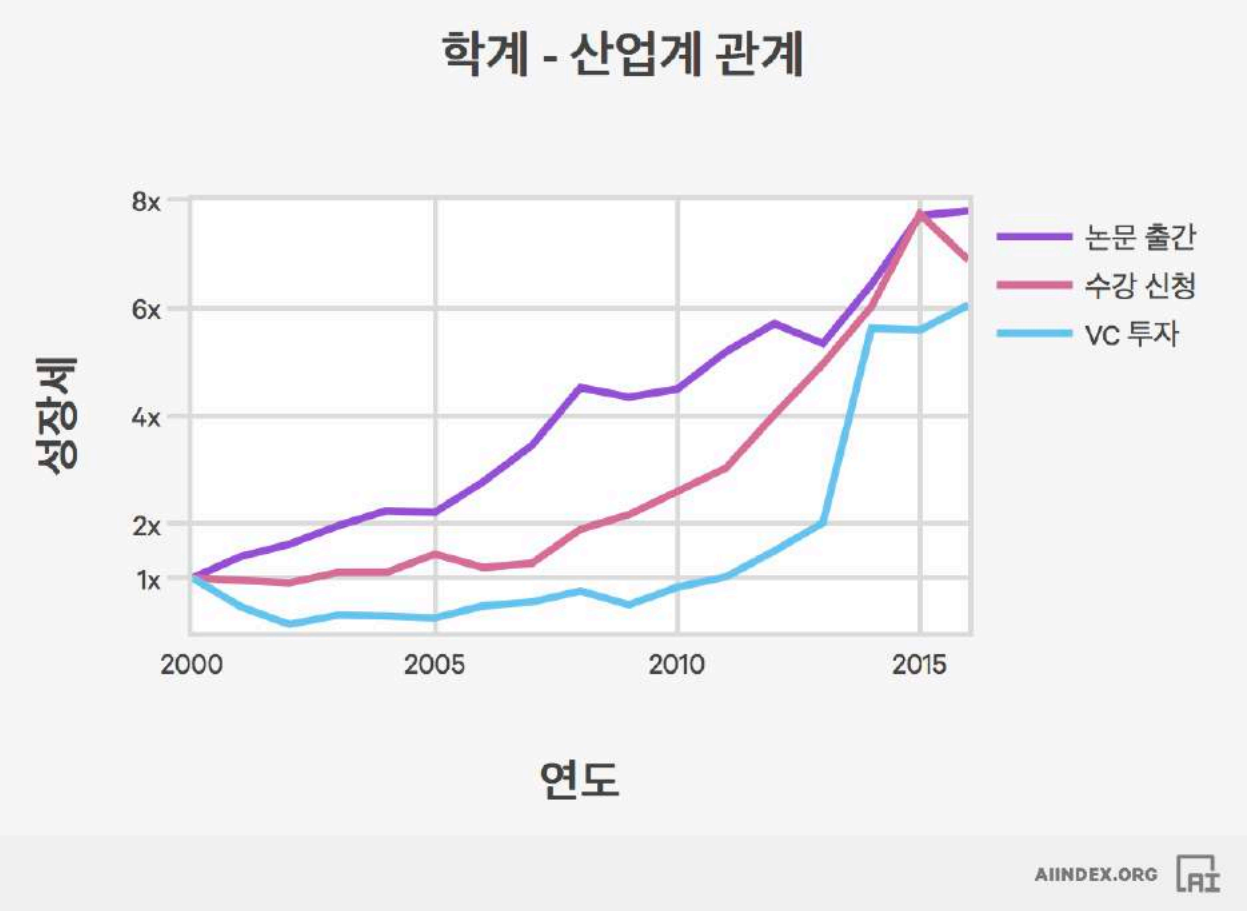
트렌드 사이의 관계 분석을 통해 앞서 살펴본 지표에서 추가적인 통찰을 얻을 수 있다. 본 섹션에서는 AI 인덱스가 수집한 데이터가 어떻게 추가 분석과 완전히 새롭고 개선된 지표의 개발을 위해 사용될 수 있는지 다룰 것이다.

이를 위한 케이스 스터디로 AI 인덱스 팀은 학계와 산업계의 트렌드 간 역학 관계를 분석하였다. 더불어 해당 지표들을 종합하여 인공지능 활성 지수(AI Vibrancy Index)를 만들었다.

학계 - 산업계 관계

학계와 산업계에서의 인공지능 관련 활동 간 관계 탐구를 위해 먼저 이전 섹션에서 다룬 몇 가지 대표적 지표를 선택하였다. 대표적인 지표로는 인공지능 관련 출간 논문 수, 스탠포드 대학교에서 인공지능과 기계학습 개론 과목 등록 학생 수와 인공지능 관련 스타트업 대상 벤처캐피털 투자를 살펴보았다.

출간 논문, 수강 신청 학생 수와 벤처캐피털 투자액은 직접 비교할 수 없는 정량적 지표다. 그러므로 해당 트렌드 간 관계 분석을 위해 각 지표를 2000년부터 시작하는 것으로 정규화(normalize)하였다. 이를 통해 지표별 절댓값 대신 지표별 증가율을 비교하였다.

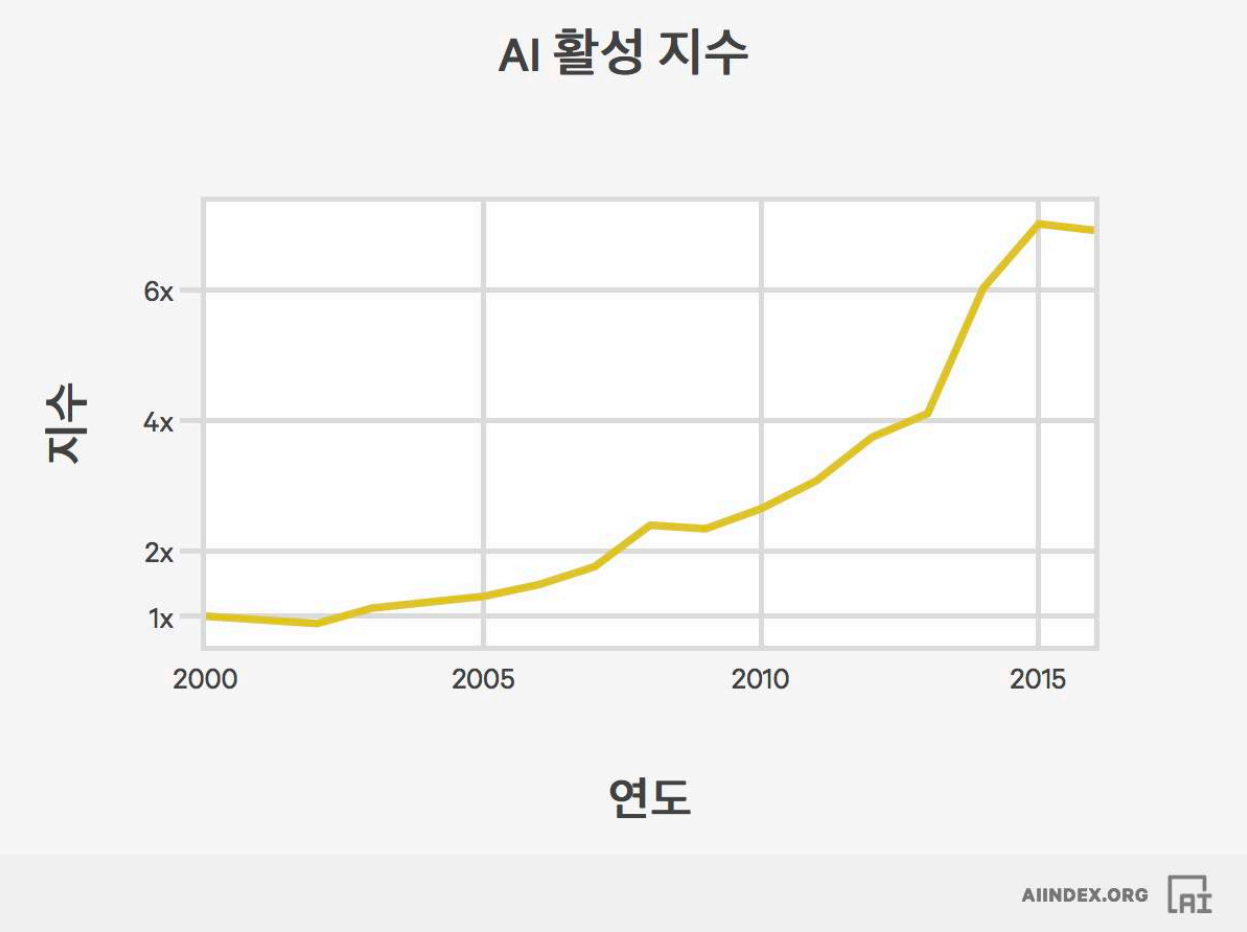


참고 : 스탠포드 대학교 *2016*학년도 인공지능, 기계학습 개론 수업 등록 학생 수 감소는 학생들의 흥미 하락으로 인한 것이 아닌 해당 학년도의 행정적 이유로 인한 것임. 자세한 내용은 부록 *A2* 참고 요망.

위 그래프에서 볼 수 있듯 처음에는 학술 활동(논문 출간 및 수강 신청)이 꾸준한 상승세를 보였다. 2010년경 투자자들이 인공지능에 주목하기 시작했으며 2013년에 이르러 투자가 총 인공지능 관련 활동의 급격한 성장을 견인했다. 그 후 학술 활동이 업계의 성장세를 따라잡는 양상을 보였다.

AI 활성 지수(The AI Vibrancy Index)

AI 활성 지수는 학계와 산업계의 지표(논문 게재, 수업 등록, 벤처캐피털 투자)를 종합하여 하나의 분야로서 인공지능의 활력(liveliness)을 정량적으로 측정한다. AI 활성 지수는 2000년도 이후 논문 출간 수, 등록 학생 수와 투자 지표를 정규화하여 평균값으로 계산하였다.



본 보고서에서 간략하게 수행한 분석이(예를 들면, 데이터의 조합으로 생성한 AI 활성 지수가), AI 인덱스에서 다룬 지표들을 이용하여 추가적으로 분석하고 어떻게 하면 보다 유용한 지표를 만들어서 추적할지에 관한 논의를 촉발하는데 도움이 되었으면 한다.

인공지능 기술과 인간 능력 비교

동일한 작업을 수행하는 인공지능 시스템과 사람의 결과물을 비교하는 것은 자연스러운 일이다. 물론 특정 작업에 있어서는 컴퓨터가 인간보다 월등히 뛰어나다. 1970년대의 휴대용 계산기가 사람보다 산술 연산을 더 잘 한다. 그러나 질문에 답하거나, 게임을 하거나, 의학적 진단을 내리는 것과 같은 보다 일반적인 작업에 있어서 인공지능 시스템의 성능을 평가하는 것은 더 어려운 일이다.

인공지능 시스템이 수행하는 작업은 구체적인 문제나 어플리케이션에서 진전을 이루기 위해 좁은 범위의 맥락에 초점이 맞추어져 있는 경우가 많다. 기계가 특정 작업을 수행하는데 뛰어난 성능을 보일 수 있지만, 만약 작업에 약간의 변화가 생길 경우 성능은 급격히 저하될 수 있다. 예를 들어 한자를 읽을 수 있는 사람은 중국어 대화도 이해하고, 중국 문화에 대해서도 아는 것이 있으며, 심지어 중국 식당에서 좋은 메뉴를 추천할 수도 있을 것이다. 이와는 대조적으로, 인공지능 시스템의 경우에는 각 경우에 맞는 매우 상이한 인공지능이 필요할 것이다.

기존 작업에 약간의 변화가 생길 경우 기계의 성능은
급격히 저하될 수 있다.

사람과 인공지능 시스템을 비교하는 것의 어려움을 감안하더라도, 컴퓨터가 인간의 수준에 도달했거나 넘어섰다고 하는 신뢰할만한 주장을 살펴보는 것은 흥미로운 일이다. 하지만 이러한 성과가 인공지능 시스템의 성능에 대해 일반화 할 수 있을 정도의 정보는 아니라는 것을 명심할 필요가 있다. 게임은 상대적으로 간단하고, 통제되고, 실험적인 환경을 제공하기 때문에 인공지능 연구에 자주 사용된다.

마일스톤

지금까지 컴퓨터로 이룬 성과와 결과물에 대해 간략하게 서술하였다. 일부 성과는 인간 수준을 향한 괄목할만한 발전을 의미하며, 일부는 초인간적 수준의 결과를 의미한다.

1980

오델로

1980년대에 리카이푸(Kai-Fu Lee)와 샌조이 마하잔(Sanjoy Mahajan)은 베이지안(Bayesian) 학습 기반 시스템을 이용한 오델로 보드 게임 프로그램 BILL을 개발하였다. 1989년 BILL은 컴퓨터 선수를 위한 전미 대회에서 우승하였으며 가장 높은 순위의 미국 선수 브라이언 로즈와의 경기에서 56대 8로 승리하였다. 1997년 로지스텔로(Logistello)라는 프로그램이 오델로 세계 챔피언과의 경기에서 6전 6승하였다.

1995

체커

1952년 아서 사무엘스(Arthur Samuels)는 자가 대국을 통해 성능을 향상시키는 일련의 체커 게임 프로그램을 만들었다. 그러나 1995년이 되어서야 체커 프로그램 치눅(Chinook)이 세계 챔피언을 제쳤다.

1997

체스

1950년대 일부 컴퓨터 과학자들은 1967년이면 컴퓨터가 인간 체스 챔피언을 능가할 것이라고 하였지만, 1997년이 되어서야 IBM의 딥블루(DeepBlue) 시스템이 체스 챔피언 게리 카스파로프(Gary Kasparov)와의 경기에서 승리했다. 오늘날 스마트폰에서 구동되는 체스 프로그램은 그랜드 마스터 수준이다.

2011

제퍼디! (Jeopardy!)

2011년 IBM 왓슨(Watson) 컴퓨터 시스템은 유명 퀴즈 쇼인 제퍼디!(Jeopardy!)[1] 에 참가하여 전 챔피언인 브레드 러터(Brad Rutter), 켄 제닝스(Ken Jennings)와의 대결에서 우승하여 100만 달러 상금을 거머쥐었다.

2015

아타리(Atari) 게임

2015년 구글 딥마인드(DeepMind) 내 팀은 강화 학습 시스템(reinforcement learning system)을 사용하여 47개의 아타리 게임을 학습시켰다. 해당 시스템은 몬테수마의 복수(Montezuma's Revenge) 등과 같은 일부 게임에선 고전을 면치 못하고 있지만, 브레이크아웃(Breakout)을 포함한 대부분의 게임에서는 인간 수준의 수행능력을 달성할 수 있게 되었다.

2016

이미지넷(ImageNet)의 객체 인식

이미지넷(ImageNet)의 객체 인식에서 자동 라벨링(labeling) 오류율(error rate)이 2010년 28%에서 2016년 3% 미만으로 감소하였다. 동일 작업에서 인간의 수행능력(인간의 오류율)은 5%이다.

[1] 제퍼디 쇼 미국의 텔레비전 퀴즈 쇼로 복잡하고 미묘한 수수께끼와 아이러니 문제 덕에 사람만이 풀 수 있다고 여겨졌었다.

2016

바둑

2016년 3월 구글 딥마인드(DeepMind)에서 개발한 알파고(AlphaGo) 시스템은 세계 최고 바둑 선수 중 한 명인 이세돌 9단과의 경기에서 4승 1패로 승리하였다. 이후 2017년 3월에 딥마인드(DeepMind)는 알파고 마스터(Alpha Go Master)를 선보였고, 세계 랭킹 1위 커제 9단을 3대 0으로 꺾었다. 2017년 10월 구글 딥마인드(DeepMind) 팀은 영국의 과학 할술지 <네이처>지에 실은 논문에서 최신 버전인 알파고 제로(AlphaGo Zero)에 대해 상세히 기술하였으며, 알파고 제로는 이세돌 9단을 4대 1로 이긴 알파고와의 대결에서 100전 100승을 거뒀다.

2017

피부암 분류(Skin Cancer Classification)

2017년 <네이처>지에 따르면 앙드레 에스테바(Andre Esteva) 등은 2,032개의 질병에 대한 129,450건의 임상 이미지를 이용하여 인공지능 시스템을 훈련시켰고, 21명의 피부과 전문의들의 진단 능력과 비교를 수행했다. 그 결과 인공지능 시스템은 피부암 진단에 있어서 피부과 전문의에 견줄 수 있는 수준을 보였다.

2017

스위치보드(Switchboard) 테스트에서의 음성인식

2017년 마이크로소프트와 IBM은 제한된 스위치보드(NIST Switchboard) 테스트에서 '인간 수준(human parity)'에 근접한 음성 인식률을 달성하였다.

2017

포커

2017년 1월 카네기멜론대학교(CMU)의 프로그램인 리브라투스(Libratus)가 네 명의 세계적 포커 선수들은 상대로 두 명이 벌이는 헤즈 업 노 리밋[2] 텍사스 홀덤(Head's Up No-Limit Texas Hold'em)이라는 포커 게임을 하였으며 총 12만 번의 승부 끝에 리브라투스의 승리로 마무리되었다. 2017년 2월 캐나다 알버타대학교의 인공지능 프로그램인 딥스택(DeepStack)은 11명의 프로 도박사들과 개인당 3천 회가 넘는 게임을 하였으며 딥스택(DeepStack)은 도박사들과의 실력 격차가 '통계학적으로 유의미한' 수준으로 우위임을 드러냈다.

2017

미스 팩맨(Ms. Pac-Man)

마이크로소프트가 인수한 딥러닝 팀인 말루바(Maluuba)는 미스 팩맨 아타리 2600 버전의 최고 점수 999,990점을 얻을 수 있는 방법을 익힌 인공지능 시스템을 개발하였다.

[2] 포커에서 헤즈 업이란 두 명이 일대일로 벌이는 게임 방식이며 노 리밋은 배팅 금액에 제한이 없다.

보고서에서 다루지 못한 내용

본 연례 보고서는 많은 내용을 다루고 있지만, 모든 것을 다루지는 못한다. 가용한 데이터 부재, 시간 부족, 또는 둘 다 부족하여 많은 중요한 부분을 보고서에서 다루지 못했다. 이하 서술한 바와 같이 본 보고서에서 다루지 못한 내용은 추후 발행될 보고서에서 다루고자 한다.

이러한 광범위한 문제를 효과적으로 해결하기 위해서는 많은 커뮤니티의 참여가 필요할 것으로 믿는다. 또한 여러분이 이하 서술한 문제를 해결할 수 있는 좋은 아이디어나 관련 데이터를 가지고 있다면 AI 인덱스 팀에 연락 부탁한다.

기술적 성능

AI 인덱스 팀은 많은 중요 기술 분야에서의 발전에 대해 다루지 못했다. 일부 분야에서는 표준화된 벤치마크가 명확하지 않았다(로봇 공학에서 대화 시스템, 계획, 연속 제어를 예로 들 수 있다). 반면, 상식적 추론(commonsense reasoning)과 같이 괄목할만한 발전이 없었던 분야에서는 성능을 측정하기가 힘들다. 또한 데이터를 수집할 기회를 갖지 못했기 때문에 아직 성과가 측정되지 않은 분야도 있다(추천 시스템, 표준화된 테스트를 예로 들 수 있다).

전통적으로 구체적인 측정이 이루어지지 않은 분야에 대한 성과 측정을 통해 인공지능의 발전에 대한 보다 냉철한 평가가 이루어질 수도 있다. 발전 정도가 지속적으로 추적되는 것은 좋은 진전이 있었을 때 이루어지는 것이 일반적이므로, 본 보고서가 과도하게 낙관적인 시각을 제시하고 있을 수도 있다.

실제로 챗봇의 대화는 인간의 대화와 비교하여 현저히 낮은 수준을 보이고 있으며 현재 이 분야에서 보편적으로 받아들여지는 벤치마크가 없는 상황이다. 유사하게, 오늘날의 인공지능 시스템은 상식적 추론 분야에서 5살 아이와 비교해서도 뒤처지는 수준이며, 상식적 추론 능력을 어떻게 정량적으로 측정해 기술적인 지표로 사용할 수 있는지도 명확하지 않다.

AI 인덱스 보고서에서 다루는 분야의 범위를 넓히는 것이 방금 언급한 낙관적 편견(optimistic bias)을 줄이는데 도움을 줄 수도 있다. 더불어 앞서 언급한 측정하기 쉽지 않은 분야에서의 효과적인 지표를 개발하려는 노력, 그 자체가 해당 분야에서 추가적 발전을 위한 원동력이 될 수 있다.

국제적 분석

전 세계적으로 일어나는 인공지능 관련 활동 및 발전이 괄목할만함에도 불구하고 본 보고서는 상당히 미국 중심적이다. 일례로 중국에서의 인공지능 관련 투자 및 활동은 놀라운 수준이지만 본 보고서에서는 다루어지지 않았다. 본 AI 인덱스 보고서를 준비하며 앞서 언급한 누락된 부분과 관련된 데이터를 수집할 역량이나 시간이 없었지만, 추후 발간될 보고서에서는 좀 더 국제적인 내용을 다룰 것이다.

다양성과 포용

인공지능 시스템 연구 및 활용은 사회에서 인공지능 기술의 영향을 결정하는데 중대한 역할을 할 것이다. 만약 우리가 인공지능이 사회에 미치는 최종적인 영향에 관한 담론에 관련 데이터를 제공하고자 한다면, 인공지능에 관한 대화에 참여하는 사람들을 정량적으로 분석하고 누가 미래의 인공지능 연구 및 활용에 영향을 미칠 수 있는 힘을 가지고 있는가를 분석해야만 한다.

본 보고서는 데이터를 생물학적 성(sex), 인종(race), 성별(gender), 민족(ethnicity), 성적 지향성(sexuality) 또는 기타 특질별로 분류하지 않았다. 얼마나 많은 여성이 산업계에서 인공지능 연구에 기여할 수 있는 기회를 갖게 되는가? 미국에서 몇 개의 인공지능 스타트업이 흑인 창업가에 의해 창업되었는가? 이러한 질문은 기술 및 벤처캐피털 산업뿐만 아니라 광범위한 시스템적 차별에 존재하는 힘의 관계와 긴밀히 관련된 질문이다. 어떠한 연구도 하나의 연구만으로는 이러한 복잡한 관계에 대한 담론의 틀을 제시할 수는 없다. 하지만

인공지능 기술이 우리 사회에 미치는 전체적인 영향에 대해 진솔한 분석을 하기 위해서는 AI 인덱스와 같은 노력이 위와 같은 질문에 답하기 위해 노력해야 한다.

정부 및 기업 투자

본 보고서에서 다룬 벤처캐피털 투자 데이터는 미국에만 초점을 맞추고 있으며, 이는 인공지능 기술 연구개발(R&D)에 대한 종합적인 투자 중 극히 일부에 불과하다.

정부와 기업에서 인공지능 연구 개발에 투자한 금액은 상당하다. 해당 투자에 관한 데이터를 구하기는 쉽지 않을 수 있지만, 미국 및 국제적 데이터 수집을 위해 노력할 것이다. 해당 분야에서 효과적인 진전을 이루기 위해서는 높은 협조 및 공동의 노력이 필요할 것으로 예상한다.

특정 분야에 미치는 영향

인공지능이 의료, 자동차, 금융, 교육 등의 분야에 미치는 영향에 대한 담론을 위해 관련 지표를 제시하고자 한다. 이러한 분야가 아마도 가장 중요하지만 가장 분석하기 어려운 분야일 것이다. 해당 분야에 관한 관련 지표는 서로 전혀 다른 분야에 대한 추가적인 이해를 요구하기 때문에 식별하고 종합하기 어렵다. AI 인덱스 팀에서는 각 분야의 여러 전문가들과 협업하여 앞서 언급한 분야로 분석을 확장해 갈 계획이다.

사회적 리스크 경감

인공지능으로 인한 사회적 리스크와 관련된 문제는 본 보고서에서 다루지 않았다. 추후에 인공지능 기술의 안전, 예측 가능성, 인공지능 알고리즘의 평등성, 인공지능 시대의 개인 정보, 자동화 증가에 따른 도덕적 영향 등에 관한 논의의 기반을 마련하는 지표를 제공하고자 한다.

전문가 논평

모든 데이터 셋 또는 인덱스는 모든 정보를 100% 담지는 못하며 의도하지 않은 편향성 (bias)을 가질 수 있다. AI 인덱스 팀이 제시하는 인덱스는 인공지능 기술의 과거, 현재, 미래에 대해 단지 부분적인 모습만을 보여줄 수 있을 뿐이다. 따라서 시야를 넓히기 위하여 학계, 업계, 정부 및 언론에서 인공지능 전문가의 식견을 듣고자 하였다.

바바라 그로스(Barbara Grosz)

하버드 대학교*(Harvard)*

보고서에서 다루지 못한 내용에 대한 이해

AI 인덱스 팀의 첫 보고서는 데이터에 기반을 둔 인공지능의 현황 분석, 짧은 시간에 보고서를 작성했다는 점과 보고서에 담지 못한 내용에 대해 AI 인덱스 팀이 유념하고 있다는 점에서 훌륭하다. 일부 누락된 내용을 보고서에 담기 위해서는 인공지능 기법의 (다른 변수를 제외한) 그 자체의 성능에 관한 지표 뿐만 아니라 개인적, 사회적 측면에서 인공지능 기술의 성능과 인간과의 상호작용 또는 인공지능 기술을 이용한 시스템이 사람들에게 미치는 영향 등을 고려한 지표가 필요할 것이다. 이러한 격차는 최근 들어 인간의 지식을 복제하는 것을 목표로 하기보다 사람의 능력을 보완하거나 증진하는 인공지능의 역량을 개발하는 것에 대한 관심이 증가한 것에 비추어 보아 특히 주목할만하다. 2016년 국제인공지능공동컨퍼런스(IJCAI-16)에서는 인간 인지 AI(human-aware AI)를 특별 주제로 다루었고, 2018년 미국인공지능학회(AAAI-18)에서는 인간과 인공지능의 협업(human-AI collaboration)을 떠오르는 주제로 다루고 있다. 또한 지난 몇 년간 다수의 워크샵이나 심포지엄에서 인간 참여 AI(human-in-the loop AI)와 AI와 사회에 관해 다루었다.

자연어 처리에 관해서도 보고서에서 다루지 못한 내용이 있다. 보고서는 파싱, 기계 번역, 특정 문서에서 질문에 대한 답을 찾는 능력에 대해 다루고 있지만 (보고서에서 밝힌 바와 같이) 대화 시스템이나 챗봇에 대해서는 다루지 않고 있다. 파싱은 분석의 대상이 되는 말(utterance)을 하는 에이전트의 내적 상태에 대한 고려를 필요로 하지 않는다. 기계 번역과 질문에 대해 응답하는 시스템 시험에 사용된 많은 환경에서도, 에이전트의 내적 상태 또는 말의 의미와 긴밀히 연관된 발화의 목적을 고려할 필요가 없다. 하지만 대화의 경우에는 상황이 다르다.

본 보고서에 언급했듯이 특정 요소가 지수에 포함되기 위해서는 이에 대한 좋은 측정법이 있어야 한다. 인공지능 기술뿐만 아니라 인간 또한 다루는 지표는 많은 도전과제를 제시하며, 임상시험 심사위원회(IRB)의 승인이 필요한 연구를 해 본 사람이라면 이에 동의할 것이다. 인공지능 기술의 알고리즘과 시스템과 관련하여 단순히 효율성이나 완성도를 넘어, 인공지능이 사람들의 삶에 미치는 영향 또한 고려하여 인공지능의 성공 여부를 판단하는 기준을 개발하는 것이 중요하다. 만약 본 보고서가 그러한 지표 개발에 박차를 가할 수 있다면 이는 인공지능, 컴퓨터 과학 그리고 사회 전반에 큰 기여가 될 것이다.

인공지능 시스템이 사람들의 삶에 미치는 영향 또한 고려하여 인공지능의 성공 여부를 판단하는 기준을 개발하는 것이 중요하다.

필자는 앞으로 발간될 AI 인덱스 보고서가 인공지능 관련 강의 수강신청 학생 수에 관해서 다룰 뿐만 아니라 인공지능이 제기하는 도덕적 문제를 다루는 수업의 수에 관해서도 다루기를 바란다(필자는 지난 3년간 “지능 시스템: 디자인 및 윤리 쟁점”이란 수업을 가르친 바 있다).

인공지능이 적용된 시스템이 우리의 일상에 점점 더 파고듦에 따라 인공지능 관련 대학 수업은 학생들에게 시스템을 디자인하는 단계에서부터 윤리적인 쟁점을 고려하는 것의 중요성에 대해 다룰 필요가 있다. AI 인덱스가 당면한 또 한가지 과제는 인공지능 시스템을 개발하는 회사 중 해당 시스템의 개발 필요성에 대해 자문해본 회사는 몇 곳이나 되는지, 인공지능 시스템 개발자 중 개발하는 시스템의 디자인이 가져올 수 있는 예상치 못한 결과와 문제를 최소화할 수 있는 방법을 고려한 개발자는 몇 명이나 되는지 등을 추적할 수 있는 방법을 찾는 것이다.

에릭 호비츠(Eric Horvitz)

마이크로소프트*(Microsoft)*

AI 인덱스의 첫 연례 보고서 발간에 대해 기쁘게 생각한다. AI 인덱스 프로젝트는 스탠포드 대학교의 인공지능 100년 연구(AI100)의 한 프로젝트로 시작하였으며, 100년에 걸쳐 인공지능 기술의 발전이 사람과 사회에 미치는 영향을 주기적으로 다루고 진단하고자 하는 AI100의 목표에 매우 부합하는 프로젝트라 할 수 있다. AI100은 인공지능과 앞으로 인공지능이 미칠 영향에 관해 인간의 지성과 주체성을 확장시켜 주는 항구적인 '연결된 실재감(connected presence)'을 정립하기 위해 설립되었다. AI 인덱스라는 야심찬 계획은 2015년 AI100 상임위원회 회의에서 처음 논의되었다.

AI 인덱스는 인공지능 기술의 발전에 관해 지속적으로 지표를 정의하고 추적한다. 첫 보고서는 인공지능의 역량, 인공지능 관련 활동에 관한 주요 지표 및 탐험적인 '파생된' 지표의 최근 트렌드에 관해 수집된 데이터를 제공하고 있다. 보고서에 제시한 지표를 통해 다방면에서 발전이 있었음을 알 수 있다. 알고리즘의 발전, 대규모 데이터 소스의 가용과 더불어, 갈수록 발전하는 컴퓨터 연산 능력에 힘입어 기계학습, 특히 컨볼루션 신경망(Convolutional Neural Networks)의 변종(variants)을 기반으로 다방면에서 최근 발전이 있었다.

파생된 지표는 대략적인 정보만 제공하며 추가 해석의 여지가 있지만, 그럼에도 불구하고 창의적이며 유용한 지표라고 생각한다. AI 활성 지수(AI Vibrancy Index)는 인공지능 기술에 관한 학계와 산업계에서의 노력과 아이디어에 관한 지표를 지속적이고 종합적으로 판단하여, 인공지능 기술의 전반적인 활성도를 측정하려는 노력의 일환이다. 추후에 발간될 보고서에서는 인공지능 관련 일자리, 기업 내 인공지능 관련 인력 구성 및 임금 통계 등 실질적인 데이터와 비교하여 검증되고 조정된 파생 지표를 내놓을 것이다.

본 보고서는 '인간 수준'의 성능에 관해 본문에서 별도의 섹션으로 다루고 있으며 상대적으로 정의 내리고 추적하기 쉬운 몇 가지 관련 성과를 예로 들고 있다. 인간 수준의 의학 진단(예를 들어 조직 절편의 시각적 분석을 통한 병리진단) 및 게임(예를 들어 오셀로, 체커, 체스, 바둑, 포커)에서 인간을 상대로 승리하는 것 등이 그 예이다. 본 보고서는 좀 더 정의하기 힘든 인간 수준의 능력에 대해서도 다루고 있다. 상식적 추론(commonsense reasoning) 관련 발전에 대해 추적하고 이해하는 것과 관련한 지표를 예로 들 수 있다. 현재의 인공지능 기술은 아기들의 상식적 이해(commonsense understanding) 수준에도 미치지 못한다.

본 보고서에서는 '보고서에서 다루지 못한 내용'에 대해서도 언급한 점이 신선하다고 생각한다. 보고서에서 명시한 누락된 부분 외에도 많은 사람들이 본 보고서의 정의나 디자인 측면에서 보완해야 할 부분, 간과하고 있는 부분, 부족한 부분을 발견할 것으로 생각한다. 그러나 AI 인덱스에 관해 논의를 시작하고 지금의 보고서를 출간해 내었다는 것만으로도 대단한 것이다. 흥미로운 지표들을 제공하는 것 외에도, 보고서 발표를 통해 인공지능 기술 발전을 이해하는데 의미 있는 기여를 하는 것은 관련 지표를 개선하고 확장하기 위한 대화에 더 넓은 커뮤니티의 참여를 촉진하는 촉매제이다.

보고서 발표를 통해 인공지능 기술 발전을 이해하는데 의미 있는 기여를 하는 것은, 관련 지표를 개선하고 확장하기 위한 대화에 더 넓은 커뮤니티의 참여를 촉진하는 촉매제이다.

시간이 지남에 따라 인공지능의 발전에 관한 지표를 제공하는 다양한 연구가 나올 것으로 기대한다. 예를 들어 AAAI 2017에서 이루어진 것처럼, 인공지능의 역량, 활동 및 영향 등의 현황과 동향에 관한 하나 혹은 그 이상의 측면을 깊게 살펴보는 것도 포함한다. 필자는 인공

지능에 관한 다양한 분석 및 지표가 나온 것을 기쁘게 생각한다. 인공지능의 영향력과 관련한 연구 및 조사가 다방면에서 증가하고 있으며 다양한 관점이 생겨날 것으로 예상할 수 있다. 그럼에도 불구하고, 좀 더 종합적인 지표를 만들기 위해 함께 노력하는 것은 가치있는 일이라고 생각한다. 해당 지표는 연구 및 조사의 중심이 될 수도 있고, 인공지능의 발전을 이해하고 추적하는데 필요한 '공통된 시각'을 제공할 수도 있다.

첫 AI 인덱스 보고서는 인공지능 발전의 최신 경향에 대해 흥미로운 통찰을 제시한다. 분석을 위한 노력이 계속 진행된다는 사실은 매우 반가운 일이며, 앞으로 인공지능 발전에 관한 그래프가 어떤 모습을 그릴지 상상하게 된다.

리 카이푸(Kai-Fu Lee)

시노베이션벤처스(創新工場, *Sinovation Ventures*)

중국의 인공지능 기술 현황

AI 인덱스는 인공지능에 관한 담론의 기반을 제공하기 위한 중요한 노력이라고 할 수 있다. 첫 보고서는 주로 미국 시장에서 인공지능 기술의 현황에 관해 이해하는데 도움이 되는 중요한 통계 자료를 많이 담고 있다. 그런 점에서 이런 (미국 이외의 인공지능 기술 현황 정보의) 부재를 메우기 위해 중국의 인공지능 기술의 현황에 대해 몇 마디 더하고자 한다.

"더 많은 데이터만큼 좋은 데이터가 없다"라는 말이 있다. 더 많은 데이터가 인공지능의 역량을 높여준다. 그렇다면 중국에서 생성되는 데이터의 양은 얼마나 될까?

중국은 전 세계에서 가장 많은 휴대폰과 인터넷 사용자를 보유하고 있으며, 이는 미국이나 인도의 휴대폰과 인터넷 사용자 수의 약 세배에 달한다. 많은 사람들은 미국과 중국의 사용자 수의 격차가 세배라고 생각하지만 실제로는 그보다 훨씬 크다. 중국에서 사람들은 휴대폰으로 결제를 하며 그 빈도는 미국에 비해 50배 더 높다. 음식 배달량의 경우 중국이 미국보다 10배 높다. 중국의 자전거 공유 서비스 모바이크(Mobike)의 경우 하루 2천만 회의 자전거 대여가 이루어지기까지 불과 10개월밖에 걸리지 않았다. 매일 2천만 대 이상의 자전거가 GPS와 기타 센서 정보를 서버로 전송하며, 이로 인해 매일 생성되는 데이터양이 20 테라바이트에 이른다. 디디추싱[3]의 경우 시범 사업으로 일부 도시의 교통 통제센터와 자사의 데이터를 연동한다고 밝힌 바 있다. 이러한 인터넷과 연결된 것들이 생산하는 데이터가 기존의 제품 및 어플리케이션을 더 효율적으로 만드는데 도움을 줄 뿐만 아니라, 우리가 미처 생각하지 못한 새로운 어플리케이션을 만들 수 있도록 할 것이다.

[3] 중국판 우버. 중국의 차량 공유 서비스로 최근 중국 우버를 인수 합병했다.

중국의 인공지능 제품의 품질은 어떠한가? 여전히 많은 사람들은 약 15년 전 중국이 그저 모조품이나 만들던 시절을 기억할 것이다. 똑똑하고 열정적인 중국의 거대 기술 기업과 창업가들은 타국 기술 기업이나 창업가들을 능가하기 위해 서구의 혁신을 기반으로 발전을 이루었다. 예를 들어, 인공지능과 관련하여 최근 중국의 안면인식 스타트업 Face++가 3개의 컴퓨터 비전 대회에서 구글, 마이크로소프트, 페이스북, 카네기멜론 대학교의 팀을 제치고 우승하는 쾌거를 이룬 바 있다.

중국 국무원은 *2030*년까지 세계 인공지능 혁신의 중심이 되겠다는 확실한 목표를 발표했다.

중국 정부는 기술 발전에 개방적이며 중국은 제품 및 서비스의 빠른 출시와 이터레이션(iteration)을 하기에 좋은 환경을 가지고 있다. 2017년 7월 중국 국무원은 『차세대 인공지능 발전 계획』을 통해 2030년까지 세계 인공지능 혁신의 중심이 되겠다는 확실한 목표를 발표했다. 중국의 계획은 인공지능을 주요 산업과 지방 정부의 우선순위 과제로 추진하는 것이다. 만약 이 계획이 말뿐인 것이라고 생각한다면 중국 정부의 계획은 과거에 고속 철도 계획, '대중창업, 만중혁신(大众创业，万众创新)' 운동[4]의 경제 정책에서 볼 수 있듯이 실제로 잘 이행되었음을 알 수 있다. 이에 미루어 인공지능 진흥 정책도 성공적으로 이행될 것으로 기대할 수 있다.

기술, 도전, 속도 친화적 속성이 중국이 인공지능 강대국이 될 수 있도록 한다. 인공지능 시대에 미국, 중국 양국이 주도권을 가지는 것이 불가피하다고 생각하며 이미 그렇게 되었다고 본다.

[4] '누구나 창업을 하고 누구나 혁신을 하자'의 뜻

앨런 맥워스(Alan Mackworth)

브리티시 콜럼비아 대학교*(University of British Columbia)*

AI 인덱스는 현재 알파 버전 단계로서 출발이 아주 좋다고 할 수 있으며 이미 인공지능의 발전을 측정하는데 유용한 도구이다. 필자는 보고서에서 다루는 내용을 확장하고 재구성하는 것과 관련하여 희망사항 몇 가지를 이야기하고자 한다. 보고서에 앞으로 추가되었으면 하는 데이터는 대부분 구하기 쉽지 않다. 그러나 가로등 효과[5]와 같은 오류를 범하지 않는 것이 중요하다. 쉽게 얻을 수 있는 데이터가 가장 유용한 데이터는 아닐 수 있다.

쉽게 얻을 수 있는 데이터가

가장 유용한 데이터는 아닐 수 있다.

가장 명백하게 본 보고서의 부족한 점으로 꼽을 수 있는 것은 대부분의 데이터가 미국 중심이라는 것이다. 그러나 미국 데이터는 얻기 쉬운 데이터이며 국제적으로 인공지능 관련 커뮤니티에서 데이터 수집에 힘을 모아주었으면 하는 바람이다. EU와 캐나다의 통계 자료가 다음으로 수집하기 쉬운 데이터일 것이다. 인공지능/기계학습 개론 수업 수강 학생 수 데이터를 예로 들 수 있다. EU에서 인공지능 연구 및 스타트업에 투자하는 금액 등도 추적 가능할 것이다. 아시아, 특히 중국 데이터는 매우 중요하며 일부는 이미 가용하다.

데이터 소스(data source)의 다양성이 부족하다는 사실을 통해 인공지능 연구자와 실무자를 지리적, 성별로 분류한 통계가 부족하다는 것을 알 수 있다.

[5] 가로등 효과란 잃어버린 열쇠를 열쇠를 잃어버린 장소에서 찾는 것이 아니라 가로등 밑이 환하기 때문에 가로등 밑에서 찾는 것과 같이 전문가들이 알고 있는 지식 범위 내에서 모든 현상을 설명하려는 것을 의미한다.

활동량 섹션에는 학계 및 산업계와 더불어 정부 또한 다루어야 할 것이다. 인공지능 연구 지원금 관련 데이터 등을 정부 활동의 예로 들 수 있으며, 다음과 같은 질문에 대해 생각해 볼 필요가 있다. 규제 기관의 활동과 거버넌스 연구를 지표로 만들 수 있는 방법이 있는가? 해당 활동 및 연구가 확실히 활발해지고 있지만, 이를 유용하게 측정할 수 있는 방법이 있는가? 또한 비정부기구도 부제로 추가하여 AI2, OpenAI, WEF와 Turing Institute와 같이 급증하는 기구에 관해서 다루는 것도 고려해 볼 필요가 있다.

학계에선 1) 매년 출간되는 인공지능/기계학습 분야 석박사 논문 수와 같은 인재 공급 데이터 및 컴퓨터 과학 논문 대비 인공지능 관련 논문 비율 2) 인공지능/기계학습 분야 교수 자릿수와 (컴퓨터 연구학회[CRA]에 게시된 박사과정 이수자 및 교수 채용 공고를 통해 알 수 있을 것이다) 같은 학계의 인력 수요 및 컴퓨터 과학 교수 대비 인공지능 교수 채용 비율에 관해 데이터를 얻을 수 있을 것이다.

많은 사람들이 관심을 가질 두 가지 지표는 산업계와 학계의 인공지능 관련직 연봉일 것이다. 헤드헌터, 컨설턴트와 CRA를 통해 유용한 자료를 얻을 수 있을 것이다. 신뢰할 수 있는 데이터를 얻기 쉬운 건 아니지만, 현재 시점에 우리가 이에 관해 알고 있는 바는 뉴욕타임즈 기사에 나오는 내용 정도에 불과하기 때문에 지표를 만들어 볼 가치가 있다.

게재 논문과 컨퍼런스 참석자 수를 학계 부제 하에 함께 다루는 것은 잘못된 분류이다. 산업계에서 이루어지는 연구 또한 활발하기 때문이다. 학계/산업계/정부/비정부 기구와 무관한 활동은 제외되어야 할 것이다. 추가적으로 유용한 지표로는 컨퍼런스 논문 제출 건수 및 아마존 카테고리에 인공지능 관련 서적으로 분류된 책 등 연간 출간되는 인공지능 관련 서적 수를 예로 들 수 있다.

기술적 성능 측면에서 인공지능이 캡차(완전 자동화된 사람과 컴퓨터 구별, CAPTCHA) 문제를 푸는데 있어 인공지능의 발전은 좋은 지표가 될 것이다. 기술적 성능 부분에서 다루지 못한 내용 중 중요한 분야가 다중 에이전트 시스템(Multi-Agent Systems, MAS)이다. 다중

에이전트 시스템 관련 활동 중 지표로 만들기 좋은 활동으로는 로보컵(RoboCup)[6]과 Trading Agent Competitions(TACs)이 있다. 일부는 뢰브너 상(Loebner Prize)[7]과 함께 튜링 테스트(Turing Test)[8]가 좋은 데이터 소스가 될 수 있다고 말한다. 이 부분에 대해서는 장단점에 대해 논의해 볼 필요가 있다. 추가해야 할 지표를 판단하는 기준에 관해 더 논의하는 것도 좋은 생각이다. (아직까지는) 정량화하기 힘든 중요 활동의 경우에는 그저 간과하지 않기 위한 목적으로 언급할 필요가 있는지도 생각해 볼 필요가 있다.

마지막으로 메타 인공지능(meta-AI) 활동 관련 지표도 생각해 봄직하다. 아직 어떻게 정량화해야 할지 명확하지는 않지만 메타 인공지능 관련 기관, 기구, 파트너십, 싱크탱크, (본 보고서와 같이) 인공지능 그 자체를 연구하는 것과 관련한 지표, 인공지능 기술 측정, 인공지능이 사회, 고용, 경제, 법, 정부, 도시에 미치는 영향 예측과 관련해 급격한 성장이 있는 것은 확실하다. 메타 인공지능의 성장과 결과에 대해 정량화하고 예측할 수 있을까? 메타-인공지능 특이점도 올까?

[6] 1996년에 시작된 연례 국제 로봇 대회로 로봇과 인공지능에 관한 연구를 위해 시작되었다.

[7] 매년 개최되는 인공지능 대회로, 가장 사람 같은 챗봇에게 수여된다. 대회의 포맷은 기본적인 튜링 테스트와 같다.

[8] 튜링 테스트는 1950년 앨런 튜링(Alan Turing)이 고안한 것으로, 인공지능을 사람과 구분할 수 있는지 테스트 하며 인공지능의 판별 기준으로 알려져 있다.

앤드류 응(Andrew Ng)

코세라, 스탠포드 대학교*(Coursera, Stanford)*

인공지능은 19세기의 전기와 같다.

인공지능은 19세기에 전기가 그러했던 것처럼 다양한 산업을 변화시키고 있다. AI 인덱스는 현 세대가 이런 사회적 변화를 추적하고 방향을 찾을 수 있도록 도와줄 것이다. 또한 미래 세대가 인공지능의 발전이 이루어진 과정을 이해하는데도 도움이 될 것이다.

인공지능은 전 세계적인 현상이다.

인공지능은 이제 전 세계적인 현상이다. AI 인덱스는 인공지능이 전 세계적으로 어떻게 발전하고 있는지 이해하기 위해서 우리 모두 각자의 국경 너머로 시야를 넓혀야 한다는 것을 상기시켜 준다. 미국과 중국은 투자 규모 및 인공지능 기술 도입 속도 면에서 선두를 달리고 있으며, 캐나다와 영국 또한 혁신적인 연구 결과를 내놓고 있다. 인공지능은 인터넷 검색에서부터 자율 주행 자동차, 고객 서비스 챗봇까지 많은 기술 시스템의 근간을 바꾸어 놓고 있다. 또한 많은 국가가 인공지능을 통해 기존에 해당 기술 시스템에서 전통적으로 앞서가던 국가를 '능가'할 수도 있다. 세심한 인공지능 관련 정책을 펴는 국가들은 더욱 빠른 속도로 발전할 것이고 충분히 심사 숙고하지 않은 정책을 펴는 국가는 뒤쳐질 위험에 처할 것이다.

세심한 인공지능 관련 정책을 펴는 국가들은

더욱 빠른 속도로 발전할 것이고,

충분히 심사 숙고하지 않은 정책을 펴는 국가는

뒤쳐질 위험에 처할 것이다.

딥러닝을 통한 인공지능 하위 분야의 변화

딥러닝은 먼저 음성인식을 변화시켰으며, 이후 컴퓨터 비전을 변화시켰다. 오늘날 자연어처리와 로봇공학이 딥러닝을 통한 혁신을 경험하고 있다. 최근 음성인식과 컴퓨터 비전 정확도의 향상은 음성인식 또는 컴퓨터 비전을 활용한 서비스의 융성으로 이어졌다. 음성인식을 활용한 서비스로는 음성인식 스피커를 예로 들 수 있으며 컴퓨터 비전을 활용한 서비스로는 자율 주행 자동차를 예로 들 수 있다. 오늘날 딥러닝을 통한 자연어처리의 혁신이 진행 중이며, 이는 챗봇과 같은 새로운 응용 프로그램의 활성화로 이어질 것이다. 로봇공학 분야에서 딥러닝의 활용 또한 괄목할만한 성장을 보이고 있으며, 이 또한 새로운 제조 역량 창출과 같은 많은 응용 사례로 이어질 것이다.

다니엘라 러스(Daniela Rus)

MIT(Massachusetts Institute of Technology)

인공지능, 긍정적인 변화로 향하는 길

세상은 빠르게 변화해왔다. 오늘날 학생들은 영상을 통해 멀리 있는 지도 교사와 만날 수 있고 의사는 수천 킬로 떨어진 곳의 환자를 영상을 통해 치료할 수 있게 되었다. 로봇은 공장에서 포장 작업을 한다. 네트워크로 연결된 센서를 통해 시설 모니터링을 할 수 있고 3D 프린팅을 이용해 맞춤형 제품을 만들 수 있다. 우리는 가능성의 세계에 살고 있다. 이러한 가능성은 우리가 인공지능과 로봇공학의 발전에 따라 무엇을 할 수 있을지 상상하면 할수록 더 커질 것이다.

전 세계를 놓고 볼 때 인공지능 기술은 우리가 직면한 가장 큰 문제를 해결하는데 있어 더 나은 통찰력을 가지도록 도와줄 것이다. 바다, 온실효과, 식물 상태를 모니터링하는 광범위한 센서 네트워크로부터 데이터를 수집, 분석함으로써 기후 변화를 이해할 수 있다. 데이터 기반 의사 결정을 통해 거버넌스를 향상시킬 수 있고, 수요와 공급의 모니터닝 및 조정을 통해 빈곤을 퇴치할 수도 있다. 가상물리 센서를 이용해 자연재해를 예측하고 대응할 수도 있다. 인공지능은 학생의 진도에 맞춰 진행되는 온라인 공개강좌(MOOC)를 통해 교육의 민주화를 이룰 수 있도록 도와줄 것이며, 모든 아이들이 좋은 직업을 구하고 행복한 삶을 사는데 필요한 역량을 기를 수 있도록 도와 줄 것이다. 아이언맨이 만화책에 나오는 등장인물이 아니라 기술적으로 가능해질 수 있는 것 처럼, 인공지능 기술은 아이들의 꿈을 현실로 만드는데도 도움을 줄 수 있다.

인공지능 기술은 우리가 직면한 가장 큰 문제들을 해결하는데 더 나은 통찰력을 가지도록 도와줄것이다.

조금 시야를 좁혀보면 인공지능은 우리의 삶을 더 안전하고 편리하고 만족스럽게 만들 수 있는 기회를 제공해 줄 것이다. 자율 주행 자동차를 타고 출퇴근하고, 자율 주행 기능 덕분에 십대 운전자의 치명적인 사고를 예방할 수 있을 것이다. 방대한 데이터를 통해 얻은 지식을 활용한 맞춤형 의료 또한 인간의 삶을 향상시킬 수 있는 기회를 제공한다. 일반적으로 생각하는 것과 반대로 인간의 삶의 향상은 만족스러운 일자리의 감소가 아니라 증가를 의미한다. 인공지능과 로봇공학으로 인한 생산성 향상이 인간을 단조로운 일에서 벗어나 컴퓨터가 할 수 없는 창의적이고 사회적이며 고난도의 일에 집중할 수 있도록 할 것이다.

우리가 인간이 기계의 도움 없이는 해결할 수 없는 문제를 해결하는데 컴퓨터를 사용한다면, 앞서 언급한 모든 것들과 그 외의 많은 것들이 가능해진다. 발전은 로봇공학, 기계학습 그리고 인공지능이라는 3개의 서로 다르지만 중첩되는 분야에서 일어나고 있다. 로봇공학은 컴퓨터 연산에 운동 능력을 제공하며 기계에 자율성을 제공한다. 인공지능 기술은 여기에 지능을 추가해 주는데, 이는 기계가 사고할 수 있게 해준다. 기계학습은 로봇공학과 인공지능을 아우르며 기계가 학습하고, 발전하고, 예측할 수 있도록 해준다. 발전은 각 분야에서 빠른 속도로 나타나고 있으며 AI 인덱스는 이러한 발전을 추적할 수 있는 몇 가지 지표를 제시한다. 해당 지표는 교육, 연구, 혁신 측면에서 인공지능 분야의 현황에 대한 중요한 정량적 분석을 제공하며, 일반적인 동향에 대한 통찰을 제공한다.

인공지능이 대단히 긍정적인 변화를 만들어낼 가능성이 있지만, 오늘날 최첨단 기술이 무엇을 할 수 있고 할 수 없는지 현황을 이해하는 것이 중요하다. AI 인덱스는 지능적 작업을 정의하고 해당 작업을 수행하는 가장 발전된 인공지능 시스템의 역량을 측정한다. 또한 AI

인덱스는 인공지능 분야와 관련한 교육과, 우리가 직면한 인공지능 관련 도전과제를 어떻게 해결할지에 대한 틀(framework)을 제공한다.

인공지능이 대단히 긍정적인 변화를 만들어낼 가능성이 있지만, 오늘날 최첨단 기술이 무엇을 할 수 있고 할 수 없는지 현황을 이해하는 것이 중요하다.

지능과 관련하여 인간의 뇌가 어떻게 지능적 활동을 수행하고, 기계가 이를 어떻게 모방할 것인가에 관한 문제는 최고의 연구자들과 장기적 연구 및 혁신이 지속되어야 하는 과학, 공학 분야의 큰 도전과제로 남아있다. AI 인덱스는 이 발전 과정을 추적하고 있는 것이다.

매간 스미스(Megan Smith)

백악관 제3대 최고기술책임자*(CTO), Shift 7*

수잔 알즈너(Susan Alzner)

UN 비정부 연락사무소*(UN Non-Governmental Liaison Service)*

AI 인덱스– 부족한 부분에서 발전을 촉진하다

2017년 연례 보고서가 제시한 목표들은 매우 중요하다. AI 인덱스는 구체적으로 "인공지능에 대해 데이터를 기반으로 한 정보 기반의 대화를 촉진하는 것을 목표"로 한다. 이는 전 세계적으로 각 커뮤니티 내에서 그리고 함께 우리가 시급히 다루어야 할 담론의 배경이 되는 새로운 환경에 대해 공유하고 지지하기 위함이다. AI 인덱스 팀은 국경을 불문하고 우리 모두가 경험하고 있는 환경과 가시적 변화를 이해하는데 있어 공백을 채우기 위해 협업하는데 앞장서서 노력하고 있다. 본 보고서에 따르면 "인공지능 기술 현황에 대해 추론할 수 있는 관련 데이터가 없다면 우리는 인공지능과 관련한 대화 및 의사 결정을 함에 있어 본질적으로 '장님 비행'을 하는 것과 마찬가지이다". '장님 비행'은 인공지능 관련분야에 직접적으로 종사하지 않는 사람, 특히 컴퓨터 과학이나 관련 기술적 배경이 없는 대부분의 사람들에게 특히 적용 가능한 표현이다.

다양성과 포용성은 무엇보다 중요하다.

우리는 편견, 차별적 문화 패턴 및 학습된 제도적 차별 행위로 인해 인류의 과반을 포용하지 못하고 있다.

본 보고서는 보고서에서 다루지 못한 내용에 관해서도 중요하게 한 섹션으로 조명하고 있으며 이와 관련하여 열린 논의의 장을 제공한다. 많은 주제가 여전히 다루어져야 하며, 논의를 시작하거나, 우선순위로 다루어야 함을 밝히고 있다. 이 중 다양성 및 포용성이 가장 중요한 부분이다. 우리는 인공지능 기술에 관해 논의하고 디자인함에 있어 인류의 과반을 놓치고 있다. 이는 거대한 의식적 또는 무의식적 편견, 심각한 차별적 문화, 거의 모든 공동체 및 매체를 불문한 언론 전반에서 목도할 수 있는 학습된 제도적 차별 행위에 기인하고 있다. 데이터를 기반으로 한 떠오르는 인공지능/기계학습 기술이 전 세계에 미치는 영향은 이미 상당하다. 우리는 수십 년에 걸쳐 거대한 변화를 목도하게 될 것이고, 심지어 앞으로 수년 안에도 엄청난 변화가 나타날 것이다. 그러므로 기술 분야의 모든 방면에서, 기술에 관한 대화에서, 관련 의사결정자들 사이에서, 모든 분야에 기술을 응용함에 있어 다양성과 포용성을 빠르고 근본적으로 향상시키는 것이 시급하다.

본 보고서에서 따르면 “인공지능 분야는 여전히 빠르게 진화하고 있으며 전문가들도 전체 영역을 아우르는 내용을 이해하거나 추적하는데 어려움을 겪고 있다.” 우리는 더 많은 사람들이 생산적인 대화에 참여할 수 있도록 유용하고 공개적인 대화의 장을 마련하고, 본 보고서와 같은 도구를 만드는데 앞장서는 전 세계의 모든 팀과 개인에게 감사한 마음을 전한다.

인공지능이 제공하는 기회와 도전과 관련하여 참고할만한 보고서는 아래와 같다.

2016년 인공지능의 미래 - 백악관 과학기술정책실(OSTP)

인공지능/기계학습에 관한 논의에 대한 폭넓은 참여의 일환으로 오바마 대통령은 과학기술정책실(OSTP) 기반 미국 정부 최고기술경영자(CTO) 팀에게 국가 과학기술 위원회(NSTC)를 포함한 정부 내 외 리더들과 함께 타운홀 미팅을 주최하라고 요청한 바 있으며, 해당 미팅은 여러 차례에 걸쳐 2016년 여름에 열렸다. 이하는 타운홀 미팅 및 2016년 10월 백악관 프런티어스 컨퍼런스에서 발표된 미국의 인공지능 계획을 포함하고 있는 보고서이다.

- 인공지능의 미래에 대한 준비 - 인공지능의 혜택과 위험에 대해 배우기 위한 워크샵 및 부처 간 실무그룹. 2016년 5월 3일, 에드 펠턴 전 백악관 과학기술정책실 차석 CTO
- 인공지능의 미래에 관한 행정부 보고서 - 인공지능의 기회, 고려 사항 및 도전. 2016년 10월 12일, 에드 펠턴 전 백악관 과학기술정책실 차석 CTO, 테라 라이온스 전 미국 최고기술경영자(CTO) 정책 자문관

참여와 응용의 확대, 윤리와 가치 통합 교육의 필수성

우리는 굉장한 전환기에 살고 있다. 인터넷은 우리 모두를 연결했고 데이터는 역대 어느 때보다 통합되고 있다. 인공지능과 기계학습의 역할이 증대됨에 따라 변화는 가속화하고 있으며, 데이터 과학, 빅 데이터, 인공지능/기계학습 및 사회가 연결됨에 따라 새로 생성되는 지능은 굉장히 중요하다. 어떻게 나쁜 결과를 예방할 수 있을까? 나쁜 결과는 스티븐 호킹이나 일론 머스크 등이 다급히 이야기하는 '로봇 재앙' 또는 대량 파괴 상황과 인공지능/기계학습이 인간의 삶 향상을 위해 제공하는 엄청난 기회(예를 들어 가난, 평등, 빈곤, 정의, 편견에 대한 대응 등과 관련한 기회)를 놓치는 것 모두 포함한다. 우리가 인공지능 활용에 관한 논의 주제를 더 빠르게 확장하거나 관련 분야에서 더 많은 사람들의 참여를 촉진하는데 속도를 낸다면, 이러한 도전을 해결하는데 도움이 될 것이다. 어찌 되었든 우리는 최대한 빠르게 공동의 가치를 관련 시스템에 반영하고 더 큰 창의성을 발휘해야 한다. 우리는 정말 인간의 행동과 관련한 데이터만을 활용하여 인공지능과 기계학습 기술을 학습시키고 싶은지 생각해 볼 필요가 있다. 인간은 선한 행동도 하지만 악한 행동 또한 행한다. 인공지능/기계학습은 이롭게 사용될 수도 있는 반면 무기로 사용될 수도 있다. 기술은 양날의 칼과 같으며 의도치 않은 용도로의 사용이든, 편견이든, 악의적인 이터레이션(malicious iterations)이든 우리가 어떻게 디자인하고 활용하느냐에 달렸다.

더 많은 사람들이 앞서 언급한 변화가 빠른 속도로 일어나고 있음을 '인지'해야 한다. 어렵더라도 윤리에 관한 담론을 나누고 변화에 대응하는 행동을 취해야 한다. 우리는 무기화된 인공지능에 대해 논의하고 우리가 도입할 수 있는 규제 및 기타 방안을 모색해야 한다. 우리가 사실이기를 바라는 것을 위하여 행동하자. 그럼에도 불구하고 다이앤 본 퍼스텐버그와 일론 머스크는 우리가 미래에 인공지능의 '애완동물'이 되는 것에 대해 이야기 할 것이다.

본 주제와 관련하여 유엔 보고서 및 세계적으로 동의한 가치에 관한 문서를 비롯해 중요한 관련 문서는 아래와 같다.

- 2030 지속 가능 개발의제(2030 Agenda for Sustainable Development & SDGs)
- 세계인권선언(Universal Declaration of Human Rights)
- 원주민 권리 선언(Declaration on Rights of Indigenous Peoples)
- 베이징 선언 및 여성 인권을 위한 행동 강령(Beijing Declaration and Platform for Action on women's human rights)
- 장애인의 권리에 관한 협약(Convention on the Rights of Persons with Disabilities)
- 경제적, 사회적 및 문화적 권리에 관한 국제규약(International Covenant on Economic, Social and Cultural Rights)
- 국제전기전자기술자협회 윤리적 디자인 일반 원칙 - 특히 인간의 혜택을 위한 원칙 1 (IEEE Ethically Aligned Design General Principles - esp Principle 1 on Human Benefit lists (p16)

2017년 MIT 졸업식 축사에서 애플 CEO 팀 쿡은 아래와 같이 말했다.

"저는 인공지능을 장착한 컴퓨터가 인간처럼 사고하는 능력을 갖게 되는 것에 대해 별로 우려하지 않습니다. 그보다 제가 오히려 더 걱정하는 건 인간이 가치나 연민과 같은 감정을 배제하고, 결과를 전혀 고려하지 않은 채 컴퓨터처럼 생각하고 행동하게 되는 상황입니다. 여러분 같은 분들이 우리가 잘못된 길로 들어서지 않게 항상 경계하고 감시해야 하는 이유이기도 합니다. 만약 과학이 어둠 속에서 무언가를 찾는 과정이라면, 인문학은 우리가 지금까지 걸어온 길과, 우리 앞에 놓인 위험에 빛을 비춰주는 촛불과도 같습니다. 스티브 잡스가 이야기했던 것처럼 기술만으로는 부족합니다. 인문학을 비롯한 다양한 순수 학문과 결합한, 우리의 가슴을 뛰게 만드는 기술이 필요합니다. 만약 여러분이 사람을 우선으로 놓고 일하신다면 그 일은 엄청난 영향력을 발휘하게 될 것입니다."

기관, 협회 및 개인이 함께 리더십을 발휘하는 것을 폭넓게 인지할 필요가 있다. 이하를 관련 예로 들 수 있다.

- 알고리즘 저스티스 리그(The Algorithmic Justice League, www.ajlunited.org/)는 알고리즘의 편견에 중점을 두며 사람들이 알고리즘의 편견에 관해 경험 및 우려를 공유하고 책임 있는 행동을 개발할 수 있는 공간을 제공함.
- 무기화된 인공지능에 대한 신속한 행동을 촉구하며 국경을 불문하고 해당 주제에 관해 참여를 촉구하는 UN 청원.
- 모두를 위한 컴퓨터 과학 프로젝트(The Computer Science for All Movement, CSforAll - csforallconsortium.org/)는 미국 및 여타 국가에서 컴퓨터 과학 관련 교육 기회 제공 확대를 위한 계획임.
- 모두를 위한 AI(The AI4All initiative, ai-4-all.org/), 새롭고 보다 다양한 인공지능 프로그래머, 학자, 리더를 양성하기 위한 계획.
- 아직 인공지능/기계학습의 활용이 많지 않은 분야의 역량 강화를 통해 인공지능/기계학습이 필요한 분야에서 긍정적인 변화를 위해 해당 기술을 활용할 수 있도록 함.
- 전통적으로 인공지능/기계학습 분야 인력을 보유하지 않은 단체/기관에서 해당 분야 전문가를 고용하도록 하기 위한 노력(예를 들어 공공 정책에 있어서 기술 지수 [TQ in Public Policy]를 활용하는 것).

오늘 발간되는 AI 인덱스의 첫 연례 보고서는 초기 단계이며 미완성에 가깝다고 할 수 있고 지금은 인지하지 못하더라도 부족한 점이 있을 것이다. 그러나 개방적 협업 방식으로 우리가 모르는 여러 가능성이 열려있는 미래로의 여정을 시작했다는 점은 대단하다 할 수 있다. 『대량 살상 수학 무기』의 저자 캐시 오닐(Cathy O'Neil)은 2017년 11월 14일 뉴욕타임즈 논평 "학계는 기술 발전에 걸맞은 교육을 제공하는 것을 계속 등한시할 순 없다("The Ivory Tower Can't Keep Ignoring Tech")에서 분야를 불문하고 학계의 참여를 독려했다. 우리는 이에 동의하며 특히 젊은 층을 포함하여 모두가 이 대화에 참여하길 바란다. AI 인덱스는 모두의 참여를 환영한다.

AI 인덱스의 첫 연례 보고서는 초기 단계이며,

미완성에 가깝다고 할 수 있고

지금은 인지하지 못하더라도 부족한 점이 있지만

시작했다는 사실만으로 대단하다 할 수 있다.

세바스찬 스런(Sebastian Thrun)

스탠포드 대학교, 유다시티*(Stanford, Udacity)*

최근 인공지능의 발전의 중요성은 아무리 강조해도 지나치지 않다. 인공지능 분야는 60년 넘게 존재해왔으며 우리에게 중대한 영향을 끼쳐왔다. 인공지능은 구글의 검색 알고리즘의 핵심이며 아마존 웹사이트 디자인과 넷플릭스 영화 추천 시스템의 바탕을 이룬다. 그러나 고성능 컴퓨터와 전례 없는 양의 데이터의 결합은 사회에 전환점이 될 것이다.

최근 몇 년간 전문직 근로자의 수준에 상응하거나 이를 능가하는 시스템이 개발되었다. 딥마인드(Deep Mind)의 알파고(AlphaGo)는 세계 최고의 바둑 선수를 꺾었다. 스탠포드 대학교 실험실에서는 인공지능 시스템이 이미지를 통해 피부과 전문의보다 더 정확하게 피부암을 진단할 수 있는 시스템을 개발하였다. 또한 필자는 구글의 자율 주행 자동차가 본인과 같은 일반 운전자보다 운전 실력이 뛰어나다고 오랫동안 주장해왔다. 현재 구글 자율주행차 팀은 일반 도로에서 안전운전 관리자(safety driver) 없이 시험운행을 하고 있다. 스타트업 크레스타(Cresta)는 인간 스페셜리스트와 인공지능 시스템이 협업하여 온라인 판매팀의 효율성을 배가할 수 있다는 것을 보여준 바 있다.

멀지 않은 미래에 인공지능이 반복적인 일로부터 인간을 해방시켜 줄 것이라고 생각한다. 인공지능 시스템은 인간 전문가가 하는 업무를 관찰하여 서서히 인간의 업무 능력을 습득할 것이다. 이로 인해 더 많은 반복적인 일이 기계의 역할이 될 것이며 인간은 보다 창의적인 일에 집중할 수 있게 될 것이다.

이러한 혁명은 역사적으로도 그 예를 찾아볼 수 있다. 증기기관의 발명 전에 인류의 대부분은 농부였다. 대부분의 사람들은 명석한 두뇌가 아니라 육체적인 힘과 민첩함으로 평가되었고

들판에서 굉장히 반복적인 일을 수행했다. 그러나 기계는 인간에게 초인적인 힘을 주었다. FarmersFeedUS.org에 따르면 미국에서 한 명의 농부가 155명분에 해당하는 곡물을 생산한다. 결과적으로 미국 인구의 2% 미만이 농업에 종사하며 98%는 다른 일을 할 수 있게 되었다. 75%의 미국인이 사무직에 종사하고 있으며 우리는 변호사, 회계사, 의사, 소프트웨어 엔지니어 등으로서 활동하고 있다. 우리가 하는 일의 대부분은 굉장히 반복적이다. 오늘날의 인공지능 기술이 반복적인 일의 패턴을 배우고 인간이 동일한 일을 더 빠르게 처리할 수 있도록 도와줄 것으로 생각한다. 궁극적으로 우리의 삶을 조직하고 세상을 이해하는데 인공지능의 도움을 받게 되면서, 우리 모두는 초인적인 힘을 갖게 될 것이다.

이러한 변화가 긍정적일까 부정적일까? 역사는 인공지능으로 인한 변화를 인류의 큰 도약으로 기록할 것으로 예상한다. 많은 사람들이 반복적인 육체노동을 하지 않게 됨으로써 우리의 교육 수준은 높아졌으며 더 창의적이 되었다. 인공지능으로 인한 새로운 혁명을 통해 우리는 전례 없는 창의성의 시대를 맞이하게 될 것으로 예상한다.

인공지능으로 인한 새로운 혁명을 통해 우리는 전례없는 창의성의 시대를 맞이하게 될 것으로 예상한다.

그러나 이러한 변화가 인간에게 지워주는 부담 또한 존재한다. 자율 주행 택시가 주 이동 수단이 된다면 일자리의 1/9이 사라질 위기에 처할 것으로 예상된다. 이러한 변화에 앞서가기 위해서 우리는 평생 배움의 자세를 가져야 한다. 새로운 역량을 키우고 기술을 습득해야 한다. 사회 전체적으로 보면 모든 사람이 변화에 적응할 수 있는 새로운 방법을 찾아야 한다.

본 보고서는 중요한 역할을 한다. 인공지능의 최근 발전을 유심히 살피며 사회에 미치는 영향을 기록한다. 신중한 조사를 통해 보고서를 작성한 저자들에게 박수를 보낸다. 본 보고서가 많은 필요성이 대두되는 인공지능에 대한 공개적 논의를 함에 있어 건설적으로 기여하길

바란다. 우리가 인공지능이 던지는 도전을 극복하고 준비하고 앞서 나가면 밝은 미래가 있을 것이다, 우리 모두에게.

마이클 울드릿지(Michael Wooldridge)

옥스포드 대학교*(University of Oxford)*

AI 인덱스 보고서는 인공지능에 기술 연구자이며, 옥스포드 대학교 컴퓨터 과학과 학과장이며 국제인공지능통합컨퍼런스(International Joint Conference on AI, IJCAI, http://www.ijcai.org/) 위원장을 역임하고, 인공지능 유럽 협회(European Association for AI, http://www.eurai.org/) 회장을 맡은 바 있는 필자에게 매우 흥미로운 보고서이다. 본 보고서는 게임, 기계 번역, 정리 증명, 질문 응답 등 다방면에서 초창기부터 인공지능 기술의 핵심적 문제로 여겨진 점을 극복하는데 있어 점진적인 개선이 이루어지고 있다는 것에 대한 종합적이고 주목할만한 자료를 제시한다. 언급한 많은 부분에서 인공지능은 이미 인간 전문가 수준에 상응하거나 이를 능가하였다. 본 보고서는 우리가 정말 증거를 필요로 했던 것 마냥 인공지능은 학생들과 산업계의 주목을 받고 있으며, 인공지능 관련 수업의 인기는 폭발하고 있고, 인공지능 관련 스타트업이 급격히 성장하고 있다는 매우 분명한 증거를 제시한다.

지금 명백히 인공지능에 관해 버블이 존재한다. 본 보고서가 제기하는 것은 버블이 (1996년-2001년 닷컴 버블처럼) 터지거나 서서히 바람이 빠져서 약해질 것인지, 그리고 그 후 무엇이 남을 것인지에 관한 질문이다. 필자가 두려워하는 것은 현재 우리가 목도하는 인공지능 기술과 관련된 대량의 투기 거품이 꺼지고, 인공지능 기술이 또 한 번의 겨울을 맞이하게 될까 하는 것이다. 현재 하고 있는 것이 무엇이든 인공지능과 관련된 것으로 속이는 것을 마다하지 않는 사기꾼과 허풍쟁이가 많이 있다. 언론에서 인공지능에 대해 좋게 말하면 잘못된 정보이고 나쁘게 말하면 광신적이라고도 할 수 있는 방송을 한다는 것에 개인적으로 굉장히 큰 좌절감을 느낀다(최근 예로 http://tinyurl.com/y9g74kkr를 들 수 있다).

현재 인공지능 기술과 관련한 버블이 앞으로 몇 년 안에 서서히 꺼지는 것이 불가피하다고 생각하지만, 버블이 극적으로 터져버리지 않고 부드럽게 가라앉을 희망이 있다고 생각한다.

이렇게 보는 가장 큰 이유는, 보고서가 명백히 보여주듯 인공지능 기술이 성과를 내고 있기 때문이다. 다양한 작업 수행에 있어 인공지능 시스템은 점진적으로 (일부는 빠르게) 성능 향상을 보이고 있으며, 이러한 강화된 인공지능의 역량은 많은 응용 분야에서 성공적으로 활용되고 있다. 달리 말하면, 필자는 현재 인공지능 버블에는 실질적인 이유가 있다고 생각하며, 이제 주요 기업들이 어떻게 인공지능 기술을 생산적으로 활용할 수 있는지 이해한다고 생각한다. 입증할 수 있는 실체가 있고 과학적으로 입증할 수 있는 발전이 있기 때문에 우리가 인공지능 기술의 겨울과 같은 역풍과 전문가 시스템 붐의 끝을 목도하게 될 것이라고 생각하지 않는다 (2027년 AI 인덱스 보고서에서 이러한 예상의 결과가 실제로 어떻게 나타날지 기대된다).

현재 인공지능 버블에는 근거가 있다고 생각한다.

본 보고서의 기술적 성능(technical performance) 부분에서 인공지능 기술에 대해 (이해할 만한 이유로) 다루지 않은 한가지 분야가 있다. 바로 보편적 인공지능(general AI)의 발전 정도이다. 해당 주제가 다루어지지 않은 이유는 필자를 포함하여 그 누구도 어떻게 보편적 인공지능의 발전 정도를 측정해야 할지 모르기 때문이다. 보고서에서 다루고 있는 주제 중 보편적 인공지능에 가장 가까운 것은 질문 응답이다. 질문 응답이 어느 정도로 인공지능의 이해력을 보여주는 것처럼 해석될 수 있지만, 완전히 보편적 인공지능이라고 하긴 힘들다. 필자는 튜링 테스트(turing test)가 그 영향력이나 독창성을 떠나서 보편적 인공지능을 측정하는 적절한 방법이라고 생각하지 않는다. 그 이유는 인공지능에 관심이 있는 독자라면 잘 알고 있을 것이다. 그렇다면 어떻게 보편적 인공지능을 향한 발전을 측정할 수 있을까? 이 질문은 보편적 인공지능의 역량이 급속도로 향상되는 것에 대한 두려움을 완화하거나 없애는 측면에서 특히 중요하다고 생각한다. 보편적 인공지능에 대한 두려움은 여전히 언론과, 짐작 건데 많은 대중의 뇌리를 떠나지 않고 있다.

참여하기

인공지능 기술의 발전과 영향을 이해하기 위한 계획이 성공하기 위해서는 다양한 커뮤니티의 참여가 필요하다고 생각한다.

다양한 커뮤니티의 참여 없이는 인공지능 기술의 영향을 이해하려는 어떠한 계획도 성공할 수 없다.

여러분이 AI 인덱스를 지원할 수 있는 여러 가지 방법이 있으며 참여를 환영한다.

AI 인덱스 2017 연례보고서에 대한 의견 공유

본 보고서에 실린 데이터에 관한 여러분의 의견을 부탁한다. 어떤 부분이 부족하며, 인공지능과 관련된 데이터를 수집하고 데이터에 관해 소통하는 것과 관련한 기회가 있다면 말씀 부탁한다. 이메일이나 트위터(@indexingai)를 통해 소중한 의견을 부탁한다.

데이터 공유

여러분이 속한 기관에서 관련 데이터를 공유할 수 있다면 연락 바란다. AI 인덱스 팀은 본 보고서 작성을 위하며 많은 기관과 협력하였으며 긴밀한 협력이 앞으로도 지속적으로 AI 인덱스의 운영 방법의 바탕이 될 것이다.

특정 분야의 지식 제공

추후 발간될 AI 인덱스 보고서는 의료, 교통, 농업 등 구체적인 분야에 인공지능이 어떠한 영향을 미치는지를 정량적으로 다룰 것이다. 이를 위해서는 여러 분야의 전문가와 협업이 필요하다. 만약 여러분이나 여러분이 아는 기관에서 인공지능이 구체적인 분야에 미치는 영향에 관한 데이터를 가지고 있다면 연락 부탁한다.

수정 사항

AI 인덱스 팀은 정확한 정보를 제공하고자 한다. 우리는 다양한 출처에서 데이터를 수집하며 이를 종합하면서 오류가 발생했을 수도 있다. 오류를 발견하셨다면 알려 주길 바란다. 웹사이트에 게시된 정보나 PDF 문서에서 수정이 필요한 부분은 수정할 것이다.

데이터 수집 지원

인공지능과 관련하여 항상 우리가 수집하고 정리할 수 있는 것보다 많은 데이터가 있을 것이다. 가장 중요하다고 할 수 있는 데이터를 수집에 있어 여러분과 협력하기를 바란다.

유용한 데이터 출처나 지표 중 인덱스 팀에서 놓치고 있는 것이 있다면 말씀 부탁한다.

국제적 정보

AI 인덱스 팀은 이미 미국을 넘어 국제적 데이터 수집을 위해 해외 파트너들과 협력하고 있다. 만약 관련된 데이터가 있다면 연락 부탁한다.

문의하기

본 보고서에서 예상치 못한 내용이 있거나 누락된 내용이 있다면 트위터 @indexingai 나 feedback@aiindex.org 로 문의 하길 바란다.

마지막으로 AI 인덱스와 인공지능의 현황에 대해 주기적으로 소식을 받고 싶으시면 aiindex.org에 이메일 소식 받기 구독 신청을 부탁한다.

감사의 말

AI 인덱스는 인덱스를 지원하고 발간을 위한 초기 자금을 제공한 스탠포드 대학교의 "인공지능에 대한 백년간 연구"(AI100) 에 감사드리며 많은 도움 주신 이하 스탠포드 대학교 관계자 여러분께 감사드린다.

Tom Abate, Amy Adams, Russ Altman, Tiffany Murray, Andrew Myers.

스타트업 자금을 지원해 준 구글, 마이크로소프트, 바이트댄스 Bytedance(터우탸오, Toutiao)에게 많은 감사를 드린다. 그러나 AI 인덱스는 해당 기관의 입장을 대변하지는 않는 독립된 활동이다.

초기 단계에 있다고 할 수 있는 AI 인덱스는 자문위원회의 식견과 제언을 통해 많은 도움을 받았다. 자문위원회 위원은 아래와 같다.

Michael Bowling, Ernie Davis, Julia Hirschberg, Eric Horvitz, Karen Levy, Alan Mackworth, Tom Mitchell, Sandy Pentland, Chris Ré, Daniela Rus, Sebastian Thrun, Hal Varian, and Toby Walsh.

전문가 논평에 참여해주신 분들께도 감사드린다. 일부는 위에 언급한 자문위원회에도 참여해주셨다. 여러분께서 제공한 식견이 AI 인덱스 보고서와 인공지능의 사회적 역할에 관한 담론을 위해 매우 유용하다.

Susan Alzner, Barbara Grosz, Eric Horvitz, Kai-Fu Lee, Alan Mackworth, Andrew Ng, Daniela Rus, Megan Smith, Sebastian Thrun, Michael Wooldridge.

귀한 제언과 도움 주신 이하 분들께도 감사드린다.

Toby Boyd, Kevin Leyton-Brown, Miles Brundage, AJ Bruno, Jeff Dean, Catherine Dong, Peter Eckersley, Stefano Ermon, Oren Etzioni, Carl Germann, Marie Hagman, Laura Hegarty, Holger Hoose, Anita Huang, Dan Jurafsky, Kevin Knight, Jure Leskovec, Tim Li, Terah Lyons, Mariano Mamertino, Christopher Manning, Gary Marcus, Dewey Murdick, Lynne Parker, Daniel Rock, Amy Sandjideh, Skyler Schain, Geoff Sutcliffe, Fabian Westerheide, Susan Woodward.

또한 앞서 언급한 분들께서 소속된 기관으로서 본 보고서를 위하여 데이터를 제공해 주신 것에 대해 감사를 표한다.

Allen Institute for Artificial Intelligence, Crunchbase, Electronic Frontier Foundation, Elsevier, EuroMatrix, Google Brain, Indeed.com, Monster.com, Sand Hill Econometrics, Sinovation Ventures, TrendKite, VentureSource.

컨퍼런스 참석자 데이터를 취득하는데 도움을 주신 분들께 감사를 표한다.

Chitta Baral, Maria Gini, Carol Hamilton, Kathryn B. Laskey, George Lee, Andrew McCallum, Laurent Michel, Mary Ellen Perry, Claude-Guy Quimper, Priscilla Rasmussen, Vesna Sabljakovic-Fritz, Terry Sejnowski, Brian Williams, Ramin Zabih.

수업 등록 학생 수 관련 정보 수집에 도움 주신 분들께 감사를 표한다.

Lance Fortnow, Charles Isbell, Leslie Kaelbling, Steven Lavalle, Dan Klein, Lenny Pitt, Mehran Saham, Tuomas Sandholm, Michael-David Sasson, Manuela Veloso, Dan Weld.

위 언급한 모든 분들의 크고 작은 도움 없이는 본 보고서를 만들 수 없었을 것이다. 도움 주신 모든 분들께 감사드리며 인공지능에 관한 담론의 기반을 만들며 지속적으로 많은 사람들의 참여를 독려하고자 한다.

부록 A: 데이터 설명 및 수집 방법

A1. 출간된 논문

출간된 논문 섹션으로 돌아가기

주요 출처 및 데이터 셋

Elsevier 출판사에서 구축한 전 세계의 우수 학술논문 인용지수 Scopus에 7천만 여종(69,794,685)의 문서가 등재되어 있다.

Scopus에 관한 추가 정보 보기.

수집된 데이터의 정의

Scopus 카탈로그에 '컴퓨터 과학(Computer Science)' 분야로 색인되어있으며 '인공지능(Artificial Intelligence)'이 주요 용어(key term)로 된 학술논문의 수.

- Scopus 데이터베이스에 '컴퓨터 과학(Computer Science)' 분야에 '인공지능(Artificial Intelligence)'이 주요 용어(key term)로 색인된 학술논문은 총 200,237권이다.
- Scopus 데이터베이스는 '컴퓨터 과학(Computer Science)' 분야에 약 5백만권(4,868,421)의 논문이 등재되어 있다.

위 논문 권수 및 Scopus 데이터베이스에 등재된 총 학술 출판물의 통계는 2017년 11월 기준이다.

데이터 수집 과정

출간된 학술 논문 데이터베이스 Scopus에 질의(query)를 만들어 인공지능과 관련된 논문, 컴퓨터 과학 분야 논문, 데이터베이스에 등재된 총 논문 수를 계산하였다. 예를 들어 2000년도 연관 논문 수를 계산하기 위하여 사용한 질의는 아래와 같다.

인공지능 관련 논문 권 수 질의

```
title-abs-key(artificial intelligence)
        AND SUBJAREA(COMP)
        AND PUBYEAR AFT 1999
        AND PUBYEAR BEF 2001
```

컴퓨터 과학 분야 논문 권 수 질의

```
SUBJAREA(COMP)
        AND PUBYEAR AFT 1999
        AND PUBYEAR BEF 2001
```

Scopus 데이터베이스에 등재된 총 논문 수 계산 질의

```
PUBYEAR AFT 1999 AND PUBYEAR BEF 2001
```

질의는 1996년부터 2016년까지 연도별로 만들었다.

Elsevier에서 Scopus API 접근권을 제공해주었기 때문에 Scopus에서 데이터를 자동으로 추출할 수 있었다.

Scopus의 질의 언어의 필드에 관한 추가 정보는 Scopus 필드 상세서 참고 요망.
Elsevier API에 관한 추가 정보는 Elsevier API 문서 참고 요망.
Scopus의 검색 API에 관한 추가 정보는 검색 API 문서 참고 요망.

뉘앙스

Scopus 시스템은 소급적으로 업데이트 된다. 결과적으로 Scopus 시스템에 등재된 논문 수는 시간이 지남에 따라 증가할 수 있다. 예를 들어 시간이 지남에 따라 Scopus에 등재되는 논문이 더 방대해짐에 따라서 "SUBJAREA (COMP) AND PUBYEAR BEF 2000" 라는 질의의 결과로 더 큰 숫자가 나올 수 있다.

Elsevier사에서는 1995년 후로 Scopus 시스템의 데이터 프로세싱이 보다 더 표준화되었기 때문에 1995년 후에 출간된 논문부터 질의가 신뢰할만한 결과값을 내놓을 것이라고 조언했다. 해당 이유로 인덱스 팀은 Scopus에서 1996년 이후로 출간된 논문에 관한 데이터만 수집하였다.

Scopus는 광범위한 출처에 기반을 둔 문헌 정보를 제공한다. Elsevier의 색인 기술 및 질의 언어 덕분에 특정 주제에 관한 논문을 찾기가 좀 더 수월하다. 출간 논문에 관한 데이터를 얻을 수 있는 출처로는 Web of Knowledge, Microsoft Academic, DBLP, CiteSeerX and Google Scholar 등이 있다. 총 논문 수는 각 데이터베이스 별로 상이할 수 있으나, 연구 논문 수 증가세가 모든 데이터베이스에서 유사한 경향을 보일 것으로 예상한다.

A2. 수강 신청

수강 신청 섹션으로 돌아가기

주요 출처 및 데이터 셋

대학교 수강 기록. 수강 관련 데이터는 이하 대학교에서 수집하였다.

University of California Berkeley, Carnegie Mellon University, Georgia Institute of Technology, University of Illinois Urbana Champaign, Massachusetts Institute of Technology, Stanford, and University of Washington.

수집된 데이터의 정의

각 학사 연도에 해당 대학교에서 대표적인 인공지능 및 기계학습 학부 수업을 수강한 학생 수. '학사 연도'는 매 해 가을에 시작한다.

데이터 수집 절차

AI 인덱스 팀은 각 대학교에서 대표적인 인공지능 및 기계학습 수업을 판별하고 학교 기록에서 수강 정보 수집을 담당하는 담당자를 통하였다.

뉘앙스

각 대학교에서 인공지능 및 기계학습에 관한 상위 수업도 제공하나 개론 수업 수강 데이터만 수집하였다. 개론 수업의 경우 모든 대학교에서 좀 더 일관되게 제공하며 분류하기 용이하다.

많은 대학교에서 인공지능 및 기계학습 수업의 수강 가능 학생 수보다 수강 수요가 더 높았다. 본 보고서의 데이터는 각 대학에서 제공해줄 수 있었던 데이터만을 기반으로 한다.

특정 연도에 특히 높은 수강생 증가율이나 하락률이 보인 것은 행정적인 이유 때문이며 학생들의 흥미 변화에 의한 것이 아니다. 예를 들어 스탠포드 대학교는 2015 -2016년 기계학습 수업 수강생 수의 하락을 이하와 같이 설명하였다.

“기계학습 수업은 보통 1년에 한번 열린다. 그러나 2015-2016년에는 (가을학기와 봄 학기에) 두 차례 열렸다. 봄 학기 수업 정보는 학생들이 가을 학기 수강 신청을 마친 뒤에야 나왔다. 그러므로 (확실한 데이터를 기반으로 하는 것은 아니지만 직감적으로) 2016-2017년도 가을학기 수업을 수강했을 학생 일부가 2015-2016년도 봄학기 수업을 수강했을 것으로 생각한다. 그로 인해 2016-2017년도 가을학기 수강생 수가 줄어든 결과를 낳았을 수 있다. 그러므로 2015-2016년 2016-2017년도 수강생 통계에 평활화(smoothing)가 필요할 것으로 보이며, 실제로 기계학습 수업에 대한 학생들의 흥미 저하가 있었다고는 보이지 않는다.”

A3. 컨퍼런스 참석자

컨퍼런스 참석자 섹션으로 돌아가기

주요 출처 및 데이터 셋

인공지능 관련 컨퍼런스 주최 측 기록. 이하 컨퍼런스 참석자 데이터를 수집하였다.

AAAI, AAMAS, ACL, CP, CVPR, ECAI, ICAPS, ICRA, ICLR, ICML, IJCAI, IROS, KR, NIPS, UAI.

수집된 데이터의 정의

인공지능 및 하위 분야와 관련한 지정된 학술 컨퍼런스 참석자 수. '대규모 컨퍼런스'는 2016년에 참석자가 1000명을 초과한 컨퍼런스로 정의하였으며 '소규모 컨퍼런스'는 2016년에 참석자가 1000명 미만인 컨퍼런스로 정의하였다.

데이터 수집 절차

AI 인덱스 팀은 각 컨퍼런스 주최 측 및 후원 기관 대표들과 함께 각 컨퍼런스 참석자 데이터를 수집하였다.

뉘앙스

모든 컨퍼런스 주최 측에서 참석자 수 데이터를 가지고 있었던 것은 아니다. 많은 관계자들이 일부 연도는 참석자 데이터가 없고 일부 연도는 대략적으로만 알고 있다고 밝혔다. 분석을 위해서는 여러 컨퍼런스 관계자들로부터 받은 대략적 통계도 정확한 것으로 간주하는 것이 적절하다고 여겼다.

모든 컨퍼런스가 매년 개최되는 것이 아니며 몇몇 컨퍼런스가 개최되지 않은 연도도 있었다.

A4. 인공지능 관련 스타트업

인공지능 관련 스타트업 섹션으로 돌아가기

주요 출처 및 데이터 셋

크런치베이스(Crunchbase)
벤쳐 소스(VentureSource), 벤처캐피털 투자를 받은 회사에 관한 종합적 데이터베이스
샌드힐 이코노메트릭스(Sand Hill Econometrics), 벤처캐피털 투자를 받은 비상장 기업에 관한 지표 제공

수집된 데이터의 정의

매년 인공지능 기술을 개발하고 있거나 활용하고 있는 것으로 밝혀진 활발히 운영되고 있는 스타트업의 수.

데이터 수집 절차

Crunchbase에서 인공지능 관련 범주에 속해있는 모든 기관 목록을 수집했다. 해당 정보 수집을 위하여 Crunchbase의 모든 범주를 검토하여 인공지능 기술 분야와 관련되었다고 여겨지는 이하 범주를 선택하였다. 범주 색인과 기관 목록은 Crunchbase가 제공한 크런치베이스 API 를 통해 얻었다.

기관 목록은 VentureSource 데이터베이스에서 얻은 모든 벤처 투자 기업 목록과 교차 대조하였다. Crunchbase 목록에 있는 벤처 투자 기업 중 VentureSource 데이터베이스에도 포함된 기업은 AI 인덱스 데이터에 포함하였다. VentureSource는 각 기업을 키워드 별로 분류하기도 하였다. VentureSource 키워드에 '인공지능'이나 '기계학습'이 포함된 기업은 관련 스타트업에 포함하였다.

Crunchbase API에 관한 추가 정보
Crunchbase 범주 목록
VentureSource의 데이터는 Sand Hill Econometrics가 담당하였다.

뉘앙스

Crunchbase에서 인공지능 관련 회사를 분류하는데 사용한 범주 목록

인공지능(Artificial Intelligence)
기계학습(Machine Learning)
자연어 처리(Natural Language Processing)
컴퓨터 비전(Computer Vision)
얼굴 인식(Facial Recognition)
이미지 인식(Image Recognition)
음성 인식(Speech Recognition)
시맨틱 검색(Semantic Search)
시맨틱 웹(Semantic Web)
텍스트 분석(Text Analytics)
가상 비서(Virtual Assistant)
시각 검색(Visual Search)
예측 분석(Predictive Analytics)
지능형 시스템(Intelligent System)

어떤 회사가 '인공지능 관련' 회사인지에 관해 간단한 답은 없다. 경험적 접근에 따르면 현재 인공지능 관련 기술로 기계학습 기술이 선호된다.

A5. 인공지능 관련 스타트업 투자

인공지능 관련 스타트업 투자 섹션으로 돌아가기

주요 출처 및 데이터 셋

크런치베이스(Crunchbase)
벤쳐 소스(VentureSource), 벤처캐피털 투자를 받은 회사에 관한 종합적 데이터베이스
샌드힐 이코노메트릭스(Sand Hill Econometrics), 벤처캐피털 투자를 받은 비장장 기업에 관한 지표 제공

수집된 데이터의 정의

인공지능이 사업 모델에서 중요한 역할을 하는 스타트업에 벤처 투자자가 매해 투자한 투자금액.

데이터 수집 절차

위 부록 A4에서 정의한 인공지능 관련 스타트업 목록을 기반으로 데이터를 수집함 (목록 보기). 지정 인공지능 관련 스타트업에 관한 투자 데이터를 VentureSource에서 검색, 종합하여 연간 투자 데이터를 만들었다.

VentureSource의 제품과의 인터렉션은 Sand Hill Econometrics가 담당하였다.

A6. 일자리

일자리 섹션으로 돌아가기

주요 출처 및 데이터 셋

인디드(Indeed.com)[9]

몬스터(Monster.com)[10]

수집된 데이터의 정의

Indeed.com 데이터는 인공지능 역량을 요구하는 각 나라별 일자리 비율을 보여주며 2013년 1월 비율을 기준으로 정규화하였다.

Monster.com 데이터는 인공지능 관련 일자리의 절댓값을 보여주며 인공지능 하위 분야의 기술을 요구하는 직업별로 분류하였다. 각 기술별로 분류된 일자리가 중복될 수도 있다. 예를 들어, 기계학습 역량을 요구하는 일자리가 자연어 처리 역량 또한 요구할 수도 있다. 이러한 일자리는 상세 그래프에서 중복 계산되었다.

데이터 수집 절차

AI 인덱스 팀은 데이터를 수집하기 위하여 Indeed사와 Monster사와 직접 협력하였다. Indeed사와 Monster사는 인공지능 관련 일자리를 분류하는 과정에서 상이한 절차를 사용하였으며 인공지능 일자리 증가율과 관련하여 다른 종류의 데이터를 제공하였다.

Indeed사는 먼저 인공지능 관련 키워드가 포함된 직책이 연관 일자리 목록에 50% 넘게 나타나는 일자리를 인공지능 관련 일자리로 분류하였다. 사용된 인공지능 관련 키워드는 아래와 같다.

---

[9] 기업 채용 게시판 및 취업정보 사이트에 게재된 수많은 취업정보를 한번에 검색해주는 검색엔진

[10] 미국 취업사이트

인공지능(Artificial Intelligence), 기계학습(Machine Learning), 자연어 처리(Natural Language Processing).

자연어 처리의 경우 인공지능 관련 키워드가 직무 설명에 포함된 일자리 중 90%가 넘는 자리와 관련되어 있었다. 인공지능 관련 직책을 조사한 후 Indeed사는 국가별로 인공지능 관련 일자리 비율을 조사하였다. Indeed사는 다년간 해당 비율을 추적하여 2013년 비율을 시작으로 정규화한(normalized) 데이터를 우리에게 송부하였다.

CEB[11]의 데이터 분석 도구 TalentNeuron이 제공한 데이터를 사용하여 Monster사는 2015, 2016, 2017(11월 10일까지) 미국에서 필수 역량에 '인공지능'을 포함한 구인광고 수를 조사하였다. 상세 분석을 위하여 동일한 TalentNeuron을 사용하여 '컴퓨터 비전'과 같은 키워드에 더해 '인공지능' 기술을 요구하는 일자리를 분류하였다.

[11] Fortune 500대 기업을 중심으로 경영 자문 서비스를 제공

A7. 로봇 수입량

로봇 수입량 섹션으로 돌아가기

주요 출처 및 데이터 셋

국제로봇 연합(International Federation of Robotics, IFR)이 발표한 세계 로보틱스 보고서(World Robotics Report).

수집된 데이터의 정의

데이터는 북미 및 국제적 연간 산업용 로봇 수입량을 보여준다. 산업용 로봇은 ISO 8373:2012 standard 기준에 정의되어 있다.

데이터 수집 절차

국제로봇 연합(International Federation of Robotics, IFR)이 발표한 세계 로보틱스 보고서(World Robotics Report)는 북미 및 국제적 연간 산업용 로봇 수입량을 보여준다. AI 인덱스 팀은 2000년부터의 수입량 데이터를 추출하였다.

뉘앙스

'인공지능'으로 분류될 수 있는 소프트웨어를 사용하는 로봇의 비율을 어떻게 계산할지는 명확하지 않다. 또한 산업용 로봇 사용에서 인공지능 기술의 기여가 어느 정도인지도 명확하지 않다.

A8. 깃허브(GitHub) 프로젝트 통계

깃허브(Github) 프로젝트 통계 섹션으로 돌아가기

주요 출처 및 데이터 셋

GitHub Archive
GitHub Archive on BigQuery

수집된 데이터의 정의

다양한 GitHub 저장소(repositories)가 받은 별의 누적 개수. 관련 repository는 이하를 포함함.

apache/incubator-mxnet, BVLC/caffe, caffe2/caffe2, dmlc/mxnet, fchollet/keras, Microsoft/CNTK, pytorch/pytorch, scikit-learn/scikit-learn, tensorflow/tensorflow, Theano/Theano.

데이터 수집 절차

GitHub archive 데이터는 Google BigQuery에 저장되어있다. AI 인덱스 팀은 구글 BigQuery에서 대상 repository의 'WatchEvents' 수를 계산하였다. 이와 관련하여 2016년 데이터를 수집하는데 사용한 코드 샘플은 아래와 같다.

```
SELECT
        project,
        YEAR(star_date) as yearly,
        MONTH(star_date) as monthly,
        SUM(daily_stars) as monthly_stars

FROM (
SELECT
        repo.name as project,
        DATE(created_at) as star_date,
        COUNT(*) as daily_stars
FROM
        TABLE_DATE_RANGE(
                [githubarchive:day.],
                TIMESTAMP("20160101"),
                TIMESTAMP("20161231"))
WHERE
        repo.name IN (
                "tensorflow/tensorflow",
                "fchollet/keras",
                "apache/incubator-mxnet",
                "scikit-learn/scikit-learn",
                "caffe2/caffe2", "pytorch/pytorch",
                "Microsoft/CNTK", "Theano/Theano",
                "dmlc/mxnet", "BVLC/caffe")
        AND type = 'WatchEvent'

GROUP BY project, star_date
)
GROUP BY project, yearly, monthly
ORDER BY project, yearly, monthly
```

뉘앙스

현재 GitHub Archive는 repository에서 사용자가 별을 제거할 경우 이를 반영할 방법은 제공하지 않고 있다. 그러므로 계산된 별의 개수는 약간 과장되었다고도 할 수 있다. 계산된 별의 개수와 실제 GitHub repository의 별 개수와 비교했을 때 숫자는 비슷하였으며 경향에도 변함이 없었다.

GitHub 별 개수를 세는 다른 방법도 존재한다. star-history 도구를 이용하여 데이터를 무작위로 검사하였다.

GitHub 프로젝트의 Fork 또한 흥미로운 조사 대상이다. 팀에서는 GitHub 저장소의 별 개수와 Fork가 거의 동일한 경향을 보이는 것을 발견하였다. Fork 데이터의 절댓값을 알고 싶다면 aiindex.org에서 데이터를 볼 수 있다. 또한 위 언급한 BigQuery 코드에 type='WatchEvent' 대신 type='ForkEvent'을 대입하여 데이터를 얻을 수도 있다.

A9. 언론 보도 동향

언론 보도 동향 섹션으로 돌아가기

주요 출처 및 데이터 셋

TrendKite: 브랜드나 에이전시의 PR이 얼마나 효과적이었는지를 측정하는 서비스를 제공하는 스타트업.

수집된 데이터의 정의

TrendKite는 언론 기사를 색인하고 기사를 '긍정적', '부정적', '중립적' 등으로 분류하여 동향을 분석한다. 인덱스 팀은 긍정적 또는 부정적으로 분류된 기사의 비율을 표시했다(나머지 기사는 중립적인 기사였다).

데이터 수집 절차

AI 인덱스 팀에서는 이하 질의를 사용하여 인공지능 관련 기사를 판별했다. 부정적인 입장을 가진 관련 없는 기사가 불균형적으로(disproportionate) 많은 매체는 제외하도록 질의를 조정했다.

Query

```
”Artificial Intelligence”
AND NOT “MarketIntelligenceCenter.com’s"

NOT site_urls_ll:(
        “individual.com"
        OR "MarketIntelligenceCenter.com")
```

TrendKite는 다양한 출처에서 기사를 가져오며 관련 검색을 할 수 있도록 필터를 제공한다. 팀은 이하 필터를 사용하였다.

- 오직 영어로 된 기사만 포함
- 보도자료 제외
- 금융 기사 제외
- 부고 제외

해당 데이터를 통해 인공지능 기술에 대한 대중의 관심 및 인공지능에 대한 언론 보도 동향을 공유하고자 하였다. 위 필터는 동향 분석 대상 기사의 범위를 단순화하기 위하여 사용하였다.

A10. 객체 인식

객체 인식 섹션으로 돌아가기

주요 출처 및 데이터 셋

LSVRC ImageNet Competition 2010 - 2017
ImageNet Data Set

수집된 데이터의 정의

2010년부터 LSVRC ImageNet 대회에서 객체 인식 부문에서 승리한 팀의 정확도를 수집하였다. 사용된 지수(Metrics)의 정의는 LSVRC 웹사이트 참고.

데이터 수집 절차

대회 관련 데이터는 ImageNet 웹사이트에서 열린 각 LSVRC 대회 리더보드에서 수집하였다.

뉘앙스

ImageNet 대회는 2017년 막을 내렸다. ILSVRC 테스트 셋에 대한 최신 결과를 보는 것이 가능할 수는 있으나 새로운 벤치마크를 식별하고 추적할 필요가 있다.

인간 수준의 능력 추정치는 Russakovsky et al, 2015를 기반으로 한다.

A11. 시각 데이터 문답(VQA)

시각 데이터 문답(VQA) 섹션으로 돌아가기

주요 출처 및 데이터 셋

Arxiv (참고문헌 검토시)
VQA Data Set

VQA 데이터 셋은 이미지, 이미지 내용에 관한 질문과 해당 질문에 대한 인간의 10가지 답변을 담고 있다.

수집된 데이터의 정의

수집된 데이터는 각 인공지능 시스템이 이미지에 대한 질문에 제한이 없는 답을 내놓는 것에 관한 정확도를 보여준다(이미지에 대한 객관식 질문에 답을 내놓는 것과 대조된다).

정확도는 VQA 원문 논문에 정의되어 있다. 인덱스 팀은 최신 결과가 나왔을 때 학술 논문을 통해 공개된 정확도를 수집했다.

데이터 수집 절차

2016년부터 2017년까지 VQA 1.0 데이터 셋에 대한 최신 결과를 발표한 논문에 대해 문헌 검토를 하였다.

뉘앙스

문헌 조사시, 최신 성과 타임라인에 약간의 변화를 주는 결과를 누락하였을 가능성이 있다. 하나의 모델이 아닌 앙상블 기법(Ensembles)이 고려되었다.

ImageNet이 컴퓨터 시각 관련 작업에서 발전을 선도하는 대회로서의 막을 내렸으므로 AI 인덱스 팀은 시각 데이터 문답 분야에서의 발전 상황을 조사하기로 하였다. 조만간 'ImageNet'을 대체할 것은 없어 보이지만, 우리가 VQA에서 한 것처럼 더 나은 벤치마크가 나오기까지 지속적으로 단절적 발전을 측정할 필요가 있다.

VQA 1.0은 데이터 셋에서 편향적인 부분을 보완하기 위해 데이터가 추가된 VQA 2.0이 나오고 얼마 지나지 않아 더 이상 사용되지 않는다.

A12. 파싱(Parsing)

파싱(parsing) 섹션으로 돌아가기

주요 출처 및 데이터 셋

Penn Treebank

Penn Treebank에서 월스트리트 저널의 문장으로 구성된 부분은 각 문장을 문장 구성 성분 분석 트리(constituency-based parse tree) 방식으로 주석을 단 데이터 셋이다. 해당 데이터의 섹션 23은 자동 파서(automatic parsers) 연구를 위한 주요 테스트 셋으로 이용하였다.

수집된 데이터의 정의

자동 파서(automatic parsers)는 자동으로 생성된 파스(parse)의 문장 구성 성분과 테스트 셋의 GOLD parse 문장 구성 성분을 비교하여 평가한다. 생성된 구성 성분의 정확도(precision)와 재현율(recall)이 반영돼 F1 점수(F1 score)가 산출된다. 보고서에 담긴 내용은 Penn Treebank의 월스트리트 저널 부분의 섹션 23에 있는 문장 파서(parser)의 F1 점수이다. 보고서에서는 F1 점수가 산정 가능한 문장들의 점수와 40글자 미만 문장의 점수를 다루었다. 문장 구성 성분 분석 트리에 관한 위키피디아 페이지를 통해 더 알아보기.

데이터 수집 절차

언제 새로운 파서(parser)가 자동 파싱의 최신 성능을 향상시켰는지 판별하기 위하여 문헌 검토를 하였다. 인덱스 팀은 1995년까지 거슬러 올라가서 발견한 파서(parser)의 F1 점수를 수집하였다. 하나의 모델이 아닌 앙상블 기법(Ensembles)이 고려되었다.

뉘앙스

자동 파싱(automatic parsers) 연구 초창기에 파서(parser)는 컴퓨터 연산 및 방법론적인 이유로 40자 미만의 문장과 100자 미만의 문장만 평가하였다. 인덱스 팀은 40자 미만의 문장과 말뭉치(corpus)에서 점수를 산정할 수 있는 모든 문장의 F1 점수를 기록하였다.

A13. 기계 번역

기계 번역 섹션으로 돌아가기

주요 출처 및 데이터 셋

컨퍼런스/기계 번역 워크샵(Workshop on Machine Translation, WMT) 뉴스 번역 부문.

EuroMatrix

연례 기계 번역 워크샵(WMT)에서 파생되어 나온 연례 기계 번역 컨퍼런스(Conference on Machine Translation)가 있다. 매해 WMT는 뉴스 번역 작업 대회를 열고 새로운 교육과 테스트 데이터 셋을 제공한다. 참가 팀은 번역 부문에 참가하기 위하여 제작한 번역 시스템을 제출한다.

수집된 데이터의 정의

WMT에서 사용하는 주요 평가 지표는 경쟁 엔트리의 순위를 매기는 것을 목표로 하며 전년도 결과와 비교하지 않는다. 또한 매우 노동 집약적이다. 인덱스 팀은 BLEU를 사용하였다, BLEU는 인간의 번역과 기계 번역의 결과의 개략적인 비교를 자동으로 수행하는 방법이다. BLEU는 변형된 버전의 정확도(precision)로 0에서 1 사이의 점수를 주며 높을수록 더 정확하다는 것을 의미한다. 기계 번역 시스템이 번역한 번역된 말뭉치(corpus)의 BLEU 평균 점수 또한 계산해 볼 수 있다. 인덱스 팀이 영어-독일어, 독일어-영어 뉴스 번역 작업을 위해 제출한 시스템이 매해 평균 BLEU 점수에서 최고점을 기록하였다. WMT 번역 작업 및 BLEU 지표의 뉘앙스에 관해 이하 참고 요망.

데이터 수집 절차

EuroMatrix는 2006년부터 제출된 영어-독일어, 독일어-영어 뉴스 번역 작업(task)의 BLEU 점수를 기록해왔다. 인덱스 팀은 각 연도마다 최고 성적을 거둔 BLEU 시스템을 선별하였다. 특히 문장을 토큰화하는(tokenizing) 프로토콜을 정의한 BLEU(11b)를 사용하는 시스템에 더 비중을 두었다. 가능하다면 영어-독일어, 독일어-영어 두 가지 경우 모두에서 높은 BLEU 점수를 얻은 시스템을 그 해 대표로 선발하였다.

11b 문장 토큰화에 관한 첨부 파일 참고 요망.

뉘앙스

BLEU는 자동으로 계산될 수 있으며 번역 품질과 인간의 판단의 상관관계를 보여줄 수 있다 여겨졌다. 그러나 BLEU 지표는 말뭉치(corpus) 간 비교에 사용될 수 없으며 시스템 간 BLEU 점수를 비교하는 것도 오해를 야기할 수 있다. 그래프는 일반적으로 우상향의 형태를 띠지만, 2016년 점수와 비교하여 2017년 BLEU 점수가 크게 떨어진 것을 통해 해당 지표가 오류가 있음을 알 수 있다(2017년 점수가 여전히 2015년 점수에 비해서는 높다). 2016년과 비교하여 기계 번역의 성능이 저하되었다기 보다 평가 시스템이 완벽하지 않을 가능성이 있다.

그러나 장기간에 걸친 경향을 보면 BLEU 점수를 통해 기계 번역 분야의 발전이 있었음을 알 수 있다.

A14. 질의응답

질의응답 섹션으로 돌아가기

주요 출처 및 데이터 셋

Stanford Question Answering Dataset

스탠포드 주최 문장 독해력 경진 데이터 셋/대회(SQuAD, The Stanford Question Answering Dataset)는 500개 이상의 기사(articles)와 기사에 대한 100,000개 이상의 질문-대답 데이터 셋이다. 기사 내용에 대한 질문이 주어지고 주어진 작업은 기사 내에서 질문에 대한 답을 찾는 것이다.

수집된 데이터의 정의

SQuAD의 성능 측정의 척도로 'Exact Match(EM)'를 사용하였다. Exact Match는 테스트 세트 내에서의 답변과 시스템이 생성한 답변이 정확히 일치하는 비율을 계산한다. 보고서에 포함된 결과는 SQuAD 데이터 셋에 대한 질의응답 시스템의 최고(state of the art) EM 점수이다.

데이터 수집 절차

SQuAD 웹사이트에 게시된 리더보드의 결과를 수집하였다.

뉘앙스

SQuAD 데이터 셋의 모든 답변은 관련 기사의 내용을 직접 인용(quotes) 한 것이다. 그러므로 시스템의 목적은 기사에서 정확히 어떤 부분에 제시된 질문에 대한 답변이 있는지 찾아내는 것이다.

추적하기는 쉽지만 얼마나 더 SQuAD 데이터 셋이 흥미를 끌지는 알 수 없다. SQuAD가 2016년 6월 발표된 후 작업(task) 수행에 대한 점수가 급격히 높아졌다.

SQuAD 데이터 셋에 대한 인간의 Exact Match 점수는 82.304로 알려졌다.

A15. 음성 인식

음성 인식 섹션으로 돌아가기

주요 출처 및 데이터 셋

Switchboard Hub5'00 데이터 셋(음성 및 전사본).

EFF AI 발전 지표

수집된 데이터의 정의

Switchboard Hub5'00 데이터 셋에 대한 학습된 음성 인식 시스템의 최신 어휘 오류율(Word Error Rate, WER) 반영. 어휘 오류율은 전사 데이터에 나타난 정답 어휘 수에 대한 인식 문장에 포함된 오류(치환, 삭제, 삽입)의 비율을 나타내는 값으로 문장 길이에 따라 정규화한 값이다. 단어 인식률(Word Accuracy), 또는 1-WER 그래프를 그림으로서 발전의 방향이 그래프 내에서 일관되게 나타나도록 하였다.

데이터 수집 절차

전자 프런티어 재단(Electronic Frontier Foundation, EFF)은 Hub5'00 데이터 셋에 대한 음성 인식 시스템의 성능을 추출하기 위하여 문헌 검토를 한 적이 있다. 본 보고서에는 해당 검토 결과를 그대로 다루었다.

뉘앙스

Switchboard 데이터는 오랫동안 사용되어 왔다. 그러므로 인공지능 시스템이 해당 데이터 셋에 과도하게 최적화(overfit) 된 것이 아닌가 하는 우려가 있으며, 해당 데이터 셋에서 발전이 앞으로 음성인식 분야의 전반적인 발전을 보여주기에는 부족한 점이 있을 수 있다.

최근 Switchboard Hub5'00 데이터 셋에 대한 인간의 WER에 대해 논쟁이 있었다. 일부는 5.1%, 일부는 5.9%, 심지어 일부 보고서는 5% 이하로 발표하였다. 본 보고서에서 5.1%를 인간의 음성인식 WER 기준으로 삼았다.

A16. 정리 증명

정리 증명 섹션으로 돌아가기

주요 출처 및 데이터 셋

Thousands of Problems for Theorem Provers (TPTP)

TPTP는 가장 큰 정리 증명 문제 인스턴스 셋이다.

수집된 데이터의 정의

자동 정리 증명(The Automatic Theorem Proving, ATP) 커뮤니티는 현재 자동 정리 증명(ATP) 기술로 풀 수 있는 문제 인스턴스의 난이도를 측정할 수 있는 방법을 개발하였다. 인덱스 팀은 TPTP 문제 일부(subset)의 평균 난이도를 기록하였고, 2010년 TPTP v5.0.0 이래로 단 한차례도 업데이트 된 적이 없는 TPTP 문제를 선택하였다. 난이도를 측정하면서 그래프의 이름을 '용이성(tractability)'으로 하였고 발전이 그래프에서 우상향의 모습을 띄도록 하였다.

TPTP 관리자(maintainer)가 작성한 TPTP 문제 난이도의 정의 문서 보기

TPTP 문제 인스턴스의 난이도가 어떻게 계산되었는지에 관한 시각적 예시

데이터 수집 절차

TPTP 데이터 셋은 각 버전에서 문제별 난이도를 포함한다. 인덱스 팀은 TPTP 문제 인스턴스에서 원하는 질문의 난이도를 추출하고 관련 문제의 평균 난이도를 지속적으로 계산하기 위하여 스크립트를 작성하였다. 스크립트는 aiindex.org에 공유할 것이다.

뉘앙스

커뮤니티에서 사용하는 TPTP 문제 난이도의 정의에는 특이한 점(quirk)이 일부 있다. 해당 난이도는 가용한 ATP 시스템에 의존적인 측면이 있다. 만약 많은 유용한 ATP 시스템이 문제를 풀지 못 할 경우 문제의 난이도는 더 어려워질 수 있다.

관리자(maintainer)가 작성한 TPTP v6.4.0 데이터셋에 관한 리뷰 보기.

A17. 충족 가능성(SAT) 문제 해결

SAT 문제 해결 섹션으로 돌아가기

주요 출처 및 데이터 셋

SAT 대회

SAT 문제해결기 데이터

충족 가능성(Satisfiability, SAT) 대회에는 현실적 성격의 문제 인스턴스를 위한 '산업' 트랙도 있다. 홀거 후스(Holger Hoos)와 케빈 레이톤 브라운(Kevin Leyton-Brown)은 2007년 부터 대회와 함께한 69개의 문제해결기(solver)와 1076개의 문제 인스턴스를 가지고 각 solver가 각 문제를 동일한 하드웨어에서 풀도록 하였다.

수집된 데이터의 정의

2007년부터 매 해 SAT 문제였던 모든 문제 중 해당 연도에 제출된 solver가 해결한 문제 비율과 최고의 solver가 해결한 문제의 비율의 평균.

데이터 수집 절차

홀거 후스(Holger Hoos)와 케빈 레이톤 브라운(Kevin Leyton-Brown)은 각 문제에 대한 각 solver의 성능을 측정하였다. AI 인덱스 팀은 단순히 해당 데이터를 취합하여 수치화하였다.

뉘앙스

홀거 후스(Holger Hoos)와 케빈 레이톤 브라운(Kevin Leyton-Brown)은 동일한 하드웨어를 사용하였지만, 이 지표는 단순히 프로세서 속도 향상으로 인해 향상될 것이다.

본 지표는 시간이 지남에 따라 SAT solver가 얼마나 효율적인지 추적하지만, 시간이 지남에 따라 새로운 SAT solver가 기여하는 바의 참신성(novelty)을 정량적으로 측정하지는 않는다. 달리 말하면 본 지표가 알고리즘의 획기적인 발전을 측정하기보다 엔지니어들의 업적을 주로 보여 줄 수도 있다. 엔지니어의 업적이 여전히 중요한 것은 사실이다. AI 인덱스 팀은 새로 만들어진 SAT solver의 근본적인 기여도를 정량적으로 측정하는 방법을 검토하고 있다.

aiindex.org