

## 123


## 오픈소스 SW R&amp;D 동향과 사례

## ① 오픈소스 R&amp;D 개요

- 오픈소스는 프리·오픈소스 소프트웨어(FOSS) 운동으로 시작되어 2000년대 기업참여 등을 통해 오픈소스 R&D로 확장
- 1980년대 리처드 스톨만이 시작한 자유소프트웨어(Free Software) 운동으로 시작되었고, 호혜적 생태계를 위한 소스코드 공개, 개방 강조
  - 1998년 넷스케이프(Netscape) 소스코드 공개 이후 개방·협력을 통한 SW개발 효율성을 확산시키기 위해 오픈소스(Open Source) 개념 도입
    - 넷스케이프와 인터넷 익스플로러간의 전쟁에서 MS의 무료화 정책으로 넷스케이프의 점유율이 하락했고 이에 대응하기 위해 오픈소스 개념을 도입하여 소스코드 공개
  - 공개된 소스코드에 대한 허용범위 등을 기준으로 자유 소프트웨어와 오픈소스 소프트웨어로 구분되어 각자의 생태계를 구축
    - 프리 소프트웨어 진영은 프리 소프트웨어 재단(Free Software Foundation)으로 대표되고, 오픈소스 진영은 아파치(Apache) 재단, 리눅스(Linux) 재단 등이 대표적
  - 오픈소스 기반의 비즈니스 모델 개발과 기업들의 기술개발-제품출시 기간 단축 노력 등에 힘입어 오픈소스 R&D로 확장
- \* 오픈소스는 정해진 사용권 범위 내에서 자유롭게 사용 하거나 변경 및 공유(배포), R&D 행위를 허용해 활용성을 강조

## &lt; 프리-오픈소스 소프트웨어 비교 &gt;

구분	주요특징
프리 소프트웨어	<ul style="list-style-type: none"> <li>◇ 프리 소프트웨어는 공유와 협업을 중시하고 이타주의를 강조하는 카피레프트(Copyleft) 캠페인으로 자유소프트웨어(Freesoftware) 재단이 대표적인 참여자</li> <li>◇ 프리 소프트웨어는 일반 공공 라이선스(GPL: General Public License)와 같이 소스 코드를 의무적으로 재공개하고 상용화를 금지하는 라이선스를 선호</li> </ul>
오픈소스 소프트웨어	<ul style="list-style-type: none"> <li>◇ 오픈소스 소프트웨어는 산업계의 발전을 위한 프로그램의 확산·보급을 통한 비즈니스에 중점을 둔 실용적인 모델로 아파치(Apache) 재단이 대표적</li> <li>◇ 오픈소스 소프트웨어 영역에서는 아파치 공공 라이선스(APL: Apache Public License) 처럼 반환 의무가 없는 라이선스를 장려</li> </ul>

 오픈소스 R&D는 개방형 혁신(Open Innovation) 방법론으로, 외부 자원을 활용하여 기존 제품의 개선, 신제품 출시가 가능

○ (경제적 이점) 신규 시장 개척을 위한 R&D 인프라 투자 절감이 가능하며, 이를 기반으로 잠재적 경쟁자 진입장벽 형성 가능

○ (이미지 차별화) 커뮤니티를 바탕으로 성장하는 비즈니스 전략을 통해 제품 및 서비스를 차별화하여 경쟁자에 대한 우위를 확보

※ 보다 혁신적인 기업 이미지 형성 등 마케팅 활용 가능

○ (품질향상) 내부자원 기반 품질 수준 한계를 외부 집단지성을 통해 극복하고, 현재까지 기술수준 검증 등 최종 제품 품질향상에 기여

○ (내부 역량강화) 오픈소스 활동을 통해 내부 개발자 역량강화, 교육비용 절감 및 역량강화 2가지 효과 동시 기대


※ 개방형 혁신(Open Innovation)으로 전체 기업문화 개선도 가능

○ (위험요소) 경쟁자 시장진입 촉진\*, 라이선스 위반\*\*, 커뮤니티 관리\*\*\*를 위한 투자비용 등 위험요소도 존재

\* 소스코드가 공개됨에 따라 기술접근이 쉬워지며, 잠재적 경쟁자 시장진입이 촉진될 가능성이 존재

\*\* 오픈소스 소스코드 활용 조건으로 영리적 목적 활용금지 등의 조항 등의 위반시 이에 대한 법적 이슈 제기 가능

\*\*\* 우수한 자원의 오픈소스 R&D 프로젝트 기여자(참여연구자)를 유지하기 위해 적절한 보상 등을 지속적으로 제공해야 함

 오픈소스 R&D는 프로젝트의 주관자, 공동 참여자들의 기여와 활용, 개발 플랫폼의 조화를 통해 수행

○ 오픈소스 R&D 프로젝트 운영을 위해서는 참여자간 역할분담과 운영 방식, 활용 라이선스 등 일정한 규칙이 존재

※ 코드 작성법, 개발 및 버그 히스토리의 관리방법, 개발된 소스의 라이선스 및 비즈니스 모델 등 포함

○ 다수의 개발자에 의해 공동으로 개발되는 프로젝트이므로, 역할은 크게 관리자(Maintainer)와 기여자(Committer)로 구분되고,

○ 개발의 방법과 배포·활용 라이선스 등은 관리자(또는 커뮤니티) 의해 제안된 프로젝트 설명서에 준하여 진행

○ 공개된 소스코드는 라이선스에 따라 활용 할 수 있으며, 특정 라이선스에 따라서는 해당 소스코드 사용 결과물의 재공개 등을 요구

- 전통적인 소프트웨어 개발과는 다른 고유의 방법론이 필요하므로, 주로 깃허브(Github)와 같은 전용 개발 플랫폼에서 R&D가 추진
  - ※ 오픈소스 프로젝트에는 어느정도까지 개발된 프로토타입을 공개하고 시작하거나 처음부터 공동으로 개발하는 방법 등 다양한 방법론이 존재
- 연구 결과물의 특허 등을 주장하기보다 공개 연구결과 플랫폼을 활용하여 재공개하거나 다수에게 공개된 방식으로 평가

< 오픈소스 R&D 용어 : Github Opensource Guide >

개념	개념설명
저장소 (Repository)	◇ 깃허브 내 오픈소스 프로젝트(Git)별 독립된 게시판
작성자 (Core Developer)	◇ 이 프로젝트를 만든 사람 혹은 조직
소유자 (Owner)	◇ 조직 또는 저장소에 대한 총 관리 권한을 가진 사람 (작성자과 일치 하지는 않음)
관리자 (Maintainer)	◇ 비전을 주도하고 프로젝트의 조직 측면을 관리하는 책임이 있는 기여자
기여자 (Committer)	◇ 프로젝트에 기여한 모든 사람
커뮤니티 멤버 (Community Members)	◇ 프로젝트를 사용하고 기여하는 사람들로 깃허브 내 의견기능을 활용하여 프로젝트 방향에 대한 의견을 표명할 수 있음
라이선스 (Licence)	◇ 해당 프로젝트의 소스코드를 활용하려 할 때 참고해야할 법적 내용을 포함한 가이드라인으로, 모든 오픈소스는 라이선스를 보유
프로젝트 설명서 (README)	◇ 새로운 커뮤니티 구성원을 위한 프로젝트 설명서로, 프로젝트 가치와 시작하는 방법을 설명
기여 가이드라인 (Contributing)	◇ 사람들이 프로젝트에 기여하는데 도움이 되도록 필요한 기여 유형과 프로세스 작동 방식을 설명
행동강령 (CODE_OF_CONDUCT)	◇ 참여자의 기본 원칙(모든 프로젝트에 존재하진 않음)
이슈 기록 (Issue Tracker)	◇ 소스코드 기여과정에서 발생했던 이슈를 시계열로 저장하는 곳
포크 (Fork)	◇ 기존 프로젝트의 개선을 위한 추가개발(Branch) 시작을 위해 원 저장소를 복사하는 행위
병합 요청 (Pull Request)	◇ 브랜치 프로젝트의 종료 시점에 다시 원 소스코드 저장소에 병합(Merge)을 요청하는 행위

## 오픈소스 기반 비즈니스 모델 개발을 통해 산업적 활용 촉진

- 오픈소스의 영리적 활용과 확산을 위해 다양한 비즈니스 모델이 개발되며, 단순 무료 소프트웨어가 아닌 비즈니스 전략으로 인식
- 업그레이드 등 사후관리를 유료 제공하는 구독모델, 상용 목적시 제품을 유료로 제공하는 듀얼 라이선스 등이 대표적
- 구독 모델, 듀얼 라이선스 외에도 지원 및 교육 모델, 호스팅 서비스 등 오픈소스 기반 영리활용을 촉진하는 다양한 비즈니스 모델이 등장

### < 오픈소스 비즈니스 모델 >

비즈니스 모델	주요내용	주요기업
구독 모델	소프트웨어를 무료로 다운로드 받고 제한 없이 사용할 수 있으나, 특정 문제를 도와 줄 컨설턴트를 고용 할 수 있으며, 기업은 업데이트 및 지원의 수준을 가입제품 기준으로 제공	RedHat, Canonical Ubuntu, CUBRID, SULinux
상업 오픈소스 제품	라이선스에 기반 하여 지원을 받는 상용 제품과 완전히 무료로 제공되는 오픈소스 제품으로 제공되는 유형	CollabNet, SugarCRM, JasperSoft, RedHat
지원 및 교육 모델	공급 업체는 하나 이상의 오픈소스 프로젝트에 지원 서비스, 교육 또는 출판물을 제공하는 유형	O'Reilly, CUBRID, 다우기술
듀얼 라이선스 모델	수요자의 니즈에 따라 GPL 또는 아파치와 같은 라이선스로 제공하기도 하고, 상용 라이선스로 제공하기도 하는 방식	MySQL, Sleepycat, Trolltech
호스팅 서비스	기업이 오픈소스 소프트웨어를 사용하는 고객에게 판매 할 수 있는 서비스를 제공	Yahoo, Google
강화된 상용SW	오픈소스 소프트웨어를 사용하고 상업적으로 제공 할 수 있는 새로운 버전은 유료 라이선스로 제공하는 방식	EnterpriseDB and SRA OSS 레드블록
컨설팅 모델	고객이 오픈소스와 관련된 전략적 의사 결정과 투자를 하는데 컨설팅을 제공하고 비용을 고객에게 청구하는 방식	IBM, Accenture, Gartner, 한국정보컨설팅
광고 모델	오픈소스 소프트웨어를 사용하고 서비스(클라우드)로 제공할 때, 해당 서비스의 사용자에게 광고를 게재하여 수익창출	Google, NHN
시스템 모델	수요자 하드웨어 제품에 오픈소스 소프트웨어를 사용할 때 계약을 기반으로 제품 판매수익을 창출	Cisco (Linksys routers), TiVo
후원 모델	돈, 장비, 또는 커뮤니티를 위한 시간 등을 제공	IBM, Google, NHN
패키지SW	한 개 이상의 오픈소스 제품을 통합하여 스택을 구성하고 공급 업체의 지원, 교육, 컨설팅 등을 함께 제공	OpenLogic, SugarCRM

## 기술 분야별로 다양한 오픈소스 솔루션이 프로젝트로 공개·운영

- OS, 스토리지 등 ▲(시스템SW), 클라우드, 웹서버 등 ▲(미들웨어), 고객관리 등 ▲(응용SW), 머신러닝 툴킷, GIS 등 ▲(기타SW)로 구분
- ※ 라이선스에 따라 완전 오픈소스 솔루션과 상용화된 오픈소스 솔루션으로 구분

### < 기술 분야별 주요 오픈소스 솔루션 현황 >

기술 분야	세부 분류	프로젝트 (솔루션명)	라이선스	주요특징
시스템 SW	운영 체제	CentOS	GPL을 포함한 Free software	Redhat Enterprise Linux(RHEL) 를 기반으로 한 오픈소스 운영체제 RHEL 수준의 기능과 확장성 제공
	스토리지	ceph	GNU LGPL v2.1	단일 분산 컴퓨터 클러스터에서 object storage 를 수행하는 free-software 스토리지 플랫폼
	가상화	KVM	GNU GPL 또는 LGPL	가상화 확장(Intel VT or AMD-V)을 제공하는 x86 하드웨어상의 리눅스에 대한 전가상화 솔루션
	데이터 관리	CUBRID	BSD, GPL2	국산 공개SW 객체 관계형 데이터베이스 관리시스템으로 데이터베이스 서버, 브로커, CUBRID 매니저로 구성되어 데이터 서비스에 최적화
	SW 공학 도구	Python	Python Foundation License	고수준 프로그래밍 언어로 C++ 또는 Java와 같은 언어를 쓰는 코드를 사용해서 표현 작고 큰 규모로 프로그램을 작성할 수 있도록 많은 라이브러리를 제공
	자원관리	Nagios	GNU GPL V2	시스템 및 네트워크 모니터링 응용프로그램
	보안	mod security	Apache V2	웹 서비스의 공격을 효과적으로 차단할 수 있는 공개 웹 응용 프로그램 방화벽 모듈 (WAF)
미들웨어	클라우드 서비스	Open stack	Apache V2	서버, 스토리지, 네트워크들과 같은 자원들을 모두 모아, 이들을 제어하고 운영하기 위한 클라우드 Operating System
	분산 시스템	Apache Camel	Apache V2	EIP(Enterprise Integration Patterns) 기반 오픈소스 통합 연계 프레임워크로 CAMEL(Concise Application Message Exchange Language)은 복잡한 라우팅 규칙을 정의하는 통합을 위한 언어
	웹 서버	Apache HTTP 서버	Apache V2	컨텐츠를 효율적으로 제공하고 (CSS, HTML) 다양한 보안 기능 (SSL, Proxy, ACL, 접근제한)을 제공하는 NCSA httpd 1.3 기반 웹서버
응용 SW	생산 도구	WordPress	GPL2	전 세계에서 가장 많이 쓰이는 세계 1위 CMS엔진 워드프레스 워드프레스는 전세계 CMS 사용량의 28%, 전세계 CMS 마켓점유율 59.5%
	CRM	Sugar CRM	Affero GPL 3	고객 중심 사업 관리를 가능하게 해주는 CRM(Customer Relationship Management) 솔루션
기타	머신러닝	Tensroflow	Apache V2	머신러닝에 필요한 기능들을 라이브러리로 제공하는 머신러닝 개발 툴킷

## ② 오픈소스 R&D 주요현황

최근 기업들은 시장 출시 소요시간(Time to Market) 감축과 빠른 시장 변화 적응을 위해 오픈소스를 도입

- 세계 65% 이상의 기업이 오픈소스를 활용 중이며, ICT 산업에서 오픈소스 소프트웨어의 비중은 갈수록 높아지는 추세

※ 삼성전자의 경우 자체 개발 제품·서비스의 90%가 오픈소스를 사용한다고 언급

- 리눅스 재단 설문조사에 따르면, 72% 기업들이 내부적·비상용 이유로 오픈소스를 사용하고, 55%는 상용 제품에 오픈소스를 활용

- 이 중 59%가 오픈소스 사용이 제품 개발 성공에 매우 중요하다 응답

- 세계 최대 오픈소스 R&D 플랫폼인 깃허브(Github)는 2018년 급격히 성장하여 패러다임 변화를 반영

- 깃허브에 따르면 전체 31백만 개발자 중 1/4에 해당하는 8백만 명이 2018년 1년 동안 신규 유입되고, 2백만 이상 단체가 깃허브 사용 시작

오픈소스 활용 뿐 아니라, 오픈소스 R&D 생태계 전체의 중요성을 인지한 글로벌 기업들의 생태계 선점 각축전도 점화

- 마이크로소프트는 세계 최대 오픈소스 R&D 플랫폼인 깃허브(Github)를 75억 달러(8조원)에 인수하고,

- IBM은 340억 달러(39조원) 규모에 리눅스(Linux, 1993)로 잘 알려진 오픈소스 소프트웨어 전문업체 레드햇(redhat) 인수

구글, MS, 삼성전자 등 국내외 선도 기업들은 단순 오픈소스 활용을 넘어 자체 신제품을 오픈소스 R&D 프로젝트로 공개

- 특히 구글, 아마존 등 글로벌 기업들은 인공지능, 클라우드 등 핵심제품을 오픈소스 R&D 프로젝트로 공개하고 배포

- 모바일 OS인 안드로이드, 머신러닝 툴킷 텐서플로우를 공개한 구글과 클라우드 가상화 도구인 AWS를 공개한 아마존 등이 대표적

- 삼성전자는 IBM, 인텔, 오라클 등과 같이 리눅스 재단의 플래티넘 등급 회원사로 적극적인 오픈소스 프로젝트의 공급자

\* 삼성은 오픈소스 R&D 전문인력 확충과 지식교류, 거버넌스 구축 등을 위해 2014년부터 삼성오픈소스 컨퍼런스(SOSCON)를 개최하고 있음

## &lt; 주요 기업들의 오픈소스 프로젝트 현황 &gt;

기업명	프로젝트명	라이선스	주요내용
Google (Alphabet)	Android	GPL v2, Apache 2.0	스마트폰을 비롯한 휴대용 장치를 위한 운영체제로 미들웨어, 사용자 인터페이스 등 포함
	Chromium	BSD, MIT 등	오픈소스 웹 브라우저 크롬 프로젝트
	Tensorflow	Apache 2.0	다양한 운영체제와 다양한 하드웨어에서 동작 가능한 머신러닝 응용프로그램을 위한 오픈소스 라이브러리
Amazon	AWS	Apache 2.0, MIT	클라우드 자원의 가상화 기술, 개발자 도구 등 100개 이상의 다양한 오픈소스 프로젝트로 구성
	ALEXA	Apache 2.0	아마존의 인공지능 음성 서비스인 알렉사를 사용하기 위한 개발자 도구를 개발하기 위한 프로젝트
IBM	APACHE	Apache 2.0	아파치 재단이 출범할 수 있었던 핵심 오픈소스 프로젝트로 1995년 처음 공개된 이후 리눅스 배포판과 함께 성장하며 HTTP 서버의 90% 이상의 점유율을 차지할 정도로 널리 보급됨
MS	vscode	MIT	비주얼 스튜디오 온라인판 기반으로 2015년 개발된 다양한 프로그래밍 언어를 지원하는 개발 환경 도구
	CNTK	MIT	대용량 데이터의 심층학습을 위한 딥러닝 프레임워크 개발을 위한 공개SW 프로젝트
Facebook	React	BSD, MIT	SNS에서 상호 작용하는 사용자 인터페이스 중심의 HTML5 앱 개발용 자바스크립트 라이브러리
Samsung	TizenRT	Apache 2.0	저사양 IoT 기기에서 작동하는 OS

## 오픈소스 R&amp;D 플랫폼의 글로벌 대기업 참여가 높은 비중 차지

- 깃허브에 따르면, 참여단체 기여도 1, 2위는 마이크로소프트(7700), 구글(5500) 등 글로벌 기업이 차지

※ 천단위에서 반올림된 수치로 저장소 마스터 브랜치에 적용된 코드만 집계한 수치

- 깃허브는 대표적인 오픈소스 R&D 플랫폼으로 6,700만개가 넘는 저장소를 보유하고, 150만개 이상의 기관(기업), 3,100만명 이상의 사용자가 활동 중

※ 대표적인 분산 버전 관리도구인 깃(Git)을 활용 오픈소스 개발을 지원하는 호스팅 서비스



- 마이크로소프트는 과거 대표적인 안티 오픈소스 기업이었으나, 오픈소스 생태계에 적극적인 참여뿐 아니라 깃허브 인수 등 태도 전환
- MS의 전 CEO인 스티브 발머는 2001년 “리눅스는 암 덩어리”라고 말하였으나, 차기 CEO 사타야 나델라는 2015년 “마이크로소프트는 리눅스를 사랑한다”고 말하여 변화된 상황을 시사

< 깃허브 참여 단체별 기여도 순위 >



상위권 기업들은 오픈소스 프로젝트를 통해 고유 개발환경 구축 추진

- MS, 페이스북, 구글 등 기업 공개 프로젝트들이 Top 10 Project 중 4개 차지
- MS의 vscode와 페이스북 react-native 및 인공지능 개발 툴킷인 구글 tensorflow 프로젝트가 1~3위
- 유사한 개발지원 툴킷인 MS Azure, Pytorch가 빠르게 성장하는 프로젝트 1, 2위로 주요 트렌드는 개발을 직관화하는 프로젝트들이 주목
- 기업들은 연구개발 진입장벽을 낮추는 클라우드 개발환경 구축을 위해 관련 오픈소스 프로젝트들을 공개

< 깃허브 주요 오픈소스 프로젝트 Top 10 >

No	Top Projects	기여자	No	Fastest growing Projects	변화*
1	Microsoft/vscode	19k	1	MicrosoftDocs/azure-docs	4.7x
2	facebook/react-native	10k	2	pytorch/pytorch	2.8x
3	tensorflow/tensorflow	9.3k	3	godotengine/godot	2.2x
4	angular/angular-cli	8.8k	4	nuxt/nuxt.js	2.1x
5	MicrosoftDocs/Azure-docs	7.8k	5	ethereum/go-ethereum	2.0x

\* 주) 2017년 기여자가 1,000명 이상인 프로젝트 중 2018년 기여(Contribution) 수의 배수



### 특히 인공지능 분야의 오픈소스 프로젝트가 급격한 성장추세

- 구글 텐서플로우 관련 저장소가 가장 많이 포크(Fork)된 프로젝트(Tensorflow, Models)에 포함
  - \* 포크는 기존 소스를 개량하기 위해 소스코드를 전부 복사하여 새로운 저장소에서 개선된 소스코드를 개발하는 행위이고, 포크 횟수가 많을수록 활발한 연구를 의미
- 텐서플로우 툴킷의 개선을 위한 Tensorflow 저장소가 가장 많이 포크된 프로젝트이며, 총 24,000여개의 Branch 프로젝트가 운영
- 툴킷을 활용한 인공지능 모델 응용 개발도 활발하여 구글 텐서플로우 모델 저장소(Repository)\*도 5위를 차지
  - \* (Tensorflow/models) 텐서플로우 기반으로 학습된 연구 결과물이 공개된 저장소

< 깃허브 활성화된 오픈소스 프로젝트 Top 5 >

No	프로젝트명	포크 프로젝트	포크 횟수
1	tensorflow	tensorflow	24K
2	twbs	bootstrap	15K
3	github	gitignore	10.8K
4	barryclark	jekyll-now	10.7K
5	tensorflow	models	8.3K


### ③ 오픈소스 R&D 주요사례 : 인공지능 분야

- 인공지능(머신러닝)은 기업참여 확대를 통해 오픈소스 R&D가 표준으로 정착한 대표적인 사례로 에코시스템이 빠르게 진화
  - 제프리 힌튼, 요슈아 벤지오, 얀 르쿤 등 리더들이 연구성과를 오픈소스로 공개하면서 인공지능 연구의 선순환 문화를 촉진
  - 또한 구글 등 강력한 산업 주체들이 경쟁적으로 오픈소스 소프트웨어 스택을 지원하며, 인공지능 연구의 선순환 구조를 지원
- 인공지능 분야의 오픈소스 에코시스템은 오픈사이언스\* 요소들을 포괄하여 4단계로 발전
  - \* 오픈사이언스란 소스코드와 데이터의 공개 외에도 성과를 공개·검증하는 오픈 액세스(Open Access) 플랫폼 등 연구협력 지원 인프라를 포괄(신은정, 2017)

- (연구개발 장벽완화) 오픈소스로 공개된 다양한 인공지능 개발 툴킷은 인공지능 R&D 진입장벽을 완화
- (프로젝트 참여촉진) 기업 참여를 통해 오픈소스 프로젝트 보상이 상금을 넘어 우수 성과자 채용연계, 연구자 네트워크 형성, 정보교류 등으로 확대되며, 연구자들의 오픈소스 프로젝트 참여가 촉진
- (연구성과 검증) 오픈소스 R&D 프로젝트 참여 결과물의 빠른 성과 검증 등을 위해 공개 학술 플랫폼(Open Access)을 활용
- (성과확산 및 선순환) 프로젝트 결과물을 오픈소스로 공개하고, 공개 연구 성과가 전이학습 등을 통해 후속 연구에 활용되는 선순환 형성

< 오픈소스 R&D 선순환 구축과정 : 인공지능 분야 >



 (연구개발 장벽완화) 오픈소스로 공개된 다양한 인공지능 개발 툴킷은 인공지능 R&D 진입장벽 완화

- 다양한 산업의 타분야 전공자도 참여가 가능하도록 인공지능 모델 개발에 필요한 기능들을 라이브러리로 제공
- 각 툴킷은 구동 환경 등 고유의 특징을 가지고 있으므로, 사용자의 상황과 목적에 따라 선택 가능

※ (ICLR 2019 사용 툴킷순위) ①텐서플로우 ②파이토치 ③케라스 ④토치 등

- 오픈소스 툴킷 자체를 오픈소스 프로젝트로 공개하여 지속 개선
  - 최근 인공지능 연구자는 툴킷의 개선을 연구하는 연구자 그룹과 툴킷을 활용하여 목적에 맞는 인공지능 모델을 개발하는 그룹으로 양분

※ 예로써 깃허브 내 Tensorflow/Tensorflow 저장소와 Tensorflow/models 저장소가 분리

- Auto ML(Machine Learning) 등 신경망 구조 설계 및 최적화 등을 추천해주는 서비스도 등장

※ 더욱 직관적이고, 전문적 지식이 적게 필요한 툴킷으로 경량화하는 추세

#### < 주요 인공지능 개발용 오픈소스 툴킷 >

소프트웨어	라이선스	오픈소스	구동환경	언어	인터페이스
Caffe (Berkely Univ)	BSD license	Yes	Linux, macOS, Windows	C++	Python, MATLAB, C++
PyTorch (Facebook, etc)	BSD license	Yes	Linux, macOS, Windows	Python, C, CUDA	Python
TensorFlow (Google)	Apache 2.0	Yes	Linux, macOS, Windows, Android	C++, Python, CUDA	Python(Keras), C/C++, Java, Go, R, Julia
Keras (François Chollet)	MIT license	Yes	Linux, macOS, Windows	Python	Python, R
Theano (Montreal Univ)	BSD license	Yes	Cross-platform	Python	Python (Keras)
Torch (Ronan Collobert et al.)	BSD license	Yes	Linux, macOS, Windows, Android, iOS	C, Lua	Lua, C, utility library for C++/OpenCL
CNTK (MS)	MIT license	Yes	Windows, Linux	C++	Python (Keras), C++

(프로젝트 참여촉진) 기업 참여를 통해 오픈소스 프로젝트 보상이 우수 성과자 채용연계, 연구 네트워크 구축 등으로 연계되어, 연구자 참여 촉진

- 구글 이미지넷(ImageNet)으로 잘 알려진 경연대회 형식의 오픈소스 프로젝트는 연구자 각자의 성취를 교류하는 계기로 활용

※ 주최 측에서 기준이 되는 데이터셋을 공개한 후, 참가자가 학습시킨 인공지능 모델의 인식률 테스트를 진행하여 상위 5개 정답률에 따라 순위 선정

※ 경연대회 결과물로 잘 알려진 시각지능 인공지능 모델인 ResNet 등이 있음

- 구글 이미지넷, 넷플릭스 프라이즈 등 글로벌 기업들도 중요성을 인식하고, 인공지능 경연대회 등 오픈소스 프로젝트를 지속 확대
  - ※ 경연대회가 다양화되며, 이를 모아서 제공하는 서비스 플랫폼도 형성 → Kaggle 등
- 캐글(Kaggle)은 구글에 자회사로 문제해결 오픈소스 프로젝트 플랫폼으로 경연대회를 통해 개발된 데이터셋과 커널(Kernels)도 개방
  - 프로젝트 참여자들의 경연대회 참가 결과는 일괄 집계되어 참여자 랭킹(Ranking)이 되고, 참여자들은 이 안에서 연구자로서 평판을 획득
- 기업 공개 프로젝트의 보상은 상금 외에도 인력 채용 루트로도 확대
  - ※ 상금은 2,500\$~1,500,000\$까지 다양하지만 주로 25,000\$~100,000\$ 규모
  - ※ 각 프로젝트에서 우수한 성과를 낸 팀(개인포함)은 채용연계 기회도 제공

< 캐글 오픈소스 프로젝트 참여 연구자 랭킹 (2019.2.15. 기준) >

133  
Grandmasters

1,116  
Masters

3,888  
Experts

45,465  
Contributors

47,317  
Novices

Rank	Tier	User		Medals	Points
1		<b>bestfitting</b>	joined 2 years ago	17  4  0	242,854
2		<b>Giba</b>	joined 6 years ago	43  32  24	171,586
3		<b>Μαρκος Μιχαηλιδης KazAnova</b>	joined 6 years ago	32  34  28	135,110
4		<b>Pavel Ostyakov</b>	joined 3 years ago	8  2  2	122,678
5		<b>Pavel Pleskov</b>	joined 4 years ago	8  15  6	119,390
6		<b>Silogram</b>	joined 6 years ago	14  28  11	118,709
7		<b>Guanshuo Xu</b>	joined 3 years ago	7  7  2	109,159
8		<b>ZFTurbo</b>	joined 3 years ago	14  19  9	108,005
9		<b>Eureka</b>	joined 5 years ago	21  21  5	98,344
10		<b>Little Boat</b>	joined 5 years ago	17  18  6	92,601

(연구성과 검증) 아카이브와 같은 공개 학술 플랫폼 활용(Open Access)을 통해 오픈소스 R&D 프로젝트 결과의 빠른 검증과 확산 추진

- 많은 연구자들은 논문을 학술지에 투고함과 동시에 아카이브(ArXiv) 등 공개 학술 플랫폼에 동시 공개하여 피어리뷰(Peer review)를 통해 성과검증
  - arXiv는 폴 긴스파그(Paul Ginsparg)가 처음 개설했고, 1991년 물리학 심사 전 논문을 보관하는 곳으로 출발했고, 천문학, 수학, 전산 과학, 비선형 과학, 계량 생물학, 통계학으로 영역을 확대

- ※ arXiv의 운영비는 현재 코넬 대학교와 미국 국립 과학 재단이 지원하고 있으며, LaTeX, PDF, DOCX 등 형식으로 제출할 수 있으며 출간물은 E-Print라고 부름
- ※ 해당 플랫폼에 참여하는 다수의 연구자들은 공개된 논문을 조회하고 이에 대해 검증
- 아카이브는 2018년에만 14만여 건의 신규 논문을 제출 받아 전년대비 14%가 성장하였고, 전세계에서 약 2억 3천만건의 다운로드 발생
- ICT에 대한 관심을 반영하여 컴퓨터 과학 분야가 전체의 24%를 차지하고, 머신러닝 등 인공지능 관련 분야는 15~25개의 논문이 매일 등록
  - ※ 2019.1.15.부터 2019.2.1.까지 인공지능 분야에 제출되어 등록된 자료 기준
- 아카이브에 등록된 논문은 위키 피어리뷰가 가능하도록 구현한 학습된 모델의 소스코드와 데이터를 함께 제출
- 아카이브를 통해 검증된 연구성과는 그 자체로 학술적 지위를 지님
- 아카이브 외에도 F1000 Research, PeerJ 등 공개 학술 플랫폼이 있음
- 연구성의 공개검증을 위한 공개 학술 플랫폼의 성장에 따라 GitXiv와 같이 성과검증에 최적화된 플랫폼도 활성화되어 확산 촉진
  - GitXiv는 arXiv에 공개한 논문과 Github에 공개한 소스코드를 동시에 보여 주는 플랫폼으로 오픈소스 프로젝트와 논문을 Rating\* 가능
  - \* 해당 분야 전문가 사이의 명성(Reputation)을 결정하는 지표로 기능
- ☐ (성과확산 및 선순환) 검증된 연구성과를 오픈소스로 개방하고, 개방된 연구성과가 2차 학습기법 등을 통해 후속 연구에 활용되는 선순환 형성
- 개발된 인공지능 모델은 전이학습(Transfer Learning) 등의 2차적 학습을 통해 전 산업분야에 필요한 특수 인공지능 모델 개발에 활용
  - ※ 인간이 가진 적응력과 유연성을 인공지능 학습에 활용할 수 있다는 점에서 발전이 기대
- 즉, 개-고양이를 구별하는 시각지능 모델은 추가 데이터 학습을 통해 암세포를 진단하는 모델이나 전혀 다른 청각 지능 모델로 변환 가능
  - 기 개발된 인공지능 모델의 가중치를 그대로 가져와서 자신이 보유한 추가 데이터를 학습(Fine Tuning)시키거나 A부터 Z까지 모든 개발을 수행(Scratch)하는 방법으로 구분
  - ※ 인간에 비유하면 눈썰미가 좋은 유관업무 경력자(높은 인식률의 모델) 소수 케이스 학습만으로도 유사업무의 숙련 노동자가 될 수 있는 것과 유사







< 인공지능 오픈소스 학습모델 저장소 사례 >


Model	TF-Slim File	Checkpoint	Top-1 Accuracy	Top-5 Accuracy
Inception V1	Code	<a href="#">inception_v1_2016_08_28.tar.gz</a>	69.8	89.6
Inception V2	Code	<a href="#">inception_v2_2016_08_28.tar.gz</a>	73.9	91.8
Inception V3	Code	<a href="#">inception_v3_2016_08_28.tar.gz</a>	78.0	93.9
Inception V4	Code	<a href="#">inception_v4_2016_09_09.tar.gz</a>	80.2	95.2
Inception-ResNet-v2	Code	<a href="#">inception_resnet_v2_2016_08_30.tar.gz</a>	80.4	95.3
ResNet V1 50	Code	<a href="#">resnet_v1_50_2016_08_28.tar.gz</a>	75.2	92.2
ResNet V1 101	Code	<a href="#">resnet_v1_101_2016_08_28.tar.gz</a>	76.4	92.9
ResNet V1 152	Code	<a href="#">resnet_v1_152_2016_08_28.tar.gz</a>	76.8	93.2
ResNet V2 50^	Code	<a href="#">resnet_v2_50_2017_04_14.tar.gz</a>	75.6	92.8
ResNet V2 101^	Code	<a href="#">resnet_v2_101_2017_04_14.tar.gz</a>	77.0	93.7
ResNet V2 152^	Code	<a href="#">resnet_v2_152_2017_04_14.tar.gz</a>	77.8	94.1
ResNet V2 200	Code	TBA	79.9*	95.2*
VGG 16	Code	<a href="#">vgg_16_2016_08_28.tar.gz</a>	71.5	89.8
VGG 19	Code	<a href="#">vgg_19_2016_08_28.tar.gz</a>	71.1	89.8
MobileNet_v1_1.0_224	Code	<a href="#">mobilenet_v1_1.0_224.tgz</a>	70.9	89.9
MobileNet_v1_0.50_160	Code	<a href="#">mobilenet_v1_0.50_160.tgz</a>	59.1	81.9
MobileNet_v1_0.25_128	Code	<a href="#">mobilenet_v1_0.25_128.tgz</a>	41.5	66.3
MobileNet_v2_1.4_224^*	Code	<a href="#">mobilenet_v2_1.4_224.tgz</a>	74.9	92.5
MobileNet_v2_1.0_224^*	Code	<a href="#">mobilenet_v2_1.0_224.tgz</a>	71.9	91.0
NASNet-A_Mobile_224#	Code	<a href="#">nasnet-a_mobile_04_10_2017.tar.gz</a>	74.0	91.6
NASNet-A_Large_331#	Code	<a href="#">nasnet-a_large_04_10_2017.tar.gz</a>	82.7	96.2
PNASNet-5_Large_331	Code	<a href="#">pnasnet-5_large_2017_12_13.tar.gz</a>	82.9	96.2
PNASNet-5_Mobile_224	Code	<a href="#">pnasnet-5_mobile_2017_12_13.tar.gz</a>	74.2	91.9

- 전이학습 관점에서 공개 모델은 새로운 인공지능 R&D의 자료이므로, 주요 공개모델을 수집·공개하는 저장소도 활성화
  - ※ 대표적인 저장소로 어썸리스트(Awesome Lists)가 알려져 있으며 Deep Learning 관련 강의, 학습 모델 등 우수 성과가 한 곳에 모여 있음
  - ※ 다수 대학도 이런 우수성과를 자발적으로 수집 공개하고 있으며, MIT 등 참여
- 이러한 저장소들을 통해 타분야 전문가도 우수 인공지능 연구성과에 손쉽게 접근 가능하며, 후속연구를 촉진

#### ④ 시사점

-  (오픈소스 R&D 활성화) 기업들의 오픈소스 R&D 참여가 확대되며, 전체 생태계가 성장하고, 특히 인공지능 분야에 관심도가 증가
  - 주요 오픈소스 R&D 프로젝트 상위 프로젝트는 기업들이 차지하고 있으며, 깃허브 등 전체 플랫폼의 양적 팽창이 확인
  - 텐서플로우 등이 인공지능 관련 프로젝트들이 상위권에 진입
-  (기술 트렌드 반영) 생태계 성장에 따라 오픈소스 프로젝트 동향을 통해 소프트웨어 전반의 트렌드 파악 가능
  - 글로벌 기업의 오픈소스 R&D 참여 비중이 상승에 따라 주요 기업들의 오픈소스 프로젝트의 동향에 전체 기술 트렌드가 반영
-  (R&D 진입장벽 완화) 전 개발 분야에서 보다 쉬운 개발 환경 조성이 촉진되며, 진입장벽이 완화될 것
  - 이런 트렌드를 개발에서 운영이 연계된다는 의미로 DevOps\*라 지칭
    - \* 개발(Development)과 운영(Operation)을 결합한 연구개발 방법론의 하나
    - DevOps 실현을 위해서는 운영자도 이해할 수 있도록 솔루션이 개발되어야 하며, 이를 위한 개발 언어, 툴킷도 보다 직관적으로 개선
    - ※ 직관적인 언어인 파이썬(Python) 및 개발자용 툴킷 등이 주요 프로젝트로 부상
-  (SW 생태계 변화) 오픈소스 R&D 활성화에 따라 소프트웨어 R&D 생태계 전반의 변화가 발생
  - 소프트웨어 개발 결과물의 배타적 권리 관계 개념이 변화되는 과정
  - 기술변화 속도가 빨라질수록 아카이브와 같은 공개 학술 플랫폼과 위키 방식의 피어리뷰 중요성은 지속 상승할 것
  - 공개 강의 등도 보편화되며 학문적 경계도 점차 약화될 전망
  - 개발자 채용역시 학위중심에서 프로젝트 성과중심으로 변화 추세
    - ※ 정규 학위과정 유무에서 오픈소스 프로젝트에서 달성한 성과를 중심으로 인재채용
  - 오픈소스 프로젝트를 통해 단절적이던 기업-연구자-타산업 연구자 간 정보 교류 및 채용 등 네트워크가 활성화



 오픈소스 R&D 확산을 통한 국내 ICT R&D 경쟁력 확보를 위해 정부 차원의 정책적 개입이 필요한 시점

\* 국내 오픈소스 생태계 참여 비중은 주요국에 비해 저조

○ 특히 공공 ICT R&D 관점에서의 정책 방향을 모색하는 후속 연구가 필요

< 깃허브 국가별 기여(Contribution) 순위 >

년도	미국	중국	인도	영국	독일	일본	한국
2014	1위	3위	5위	2위	4위	8위	19위
2016	1위	2위	3위	4위	5위	10위	16위
2018	1위	2위	3위	4위	5위	8위	16위