

# 토픽 모델링을 이용한 인공지능(A.I.) 기술 동향 예측

서울과학기술대학교 일반대학원 데이터사이언스학과  
20510082 전우진



# 목차

1. 주제

2. 데이터

- 수집 및 전처리

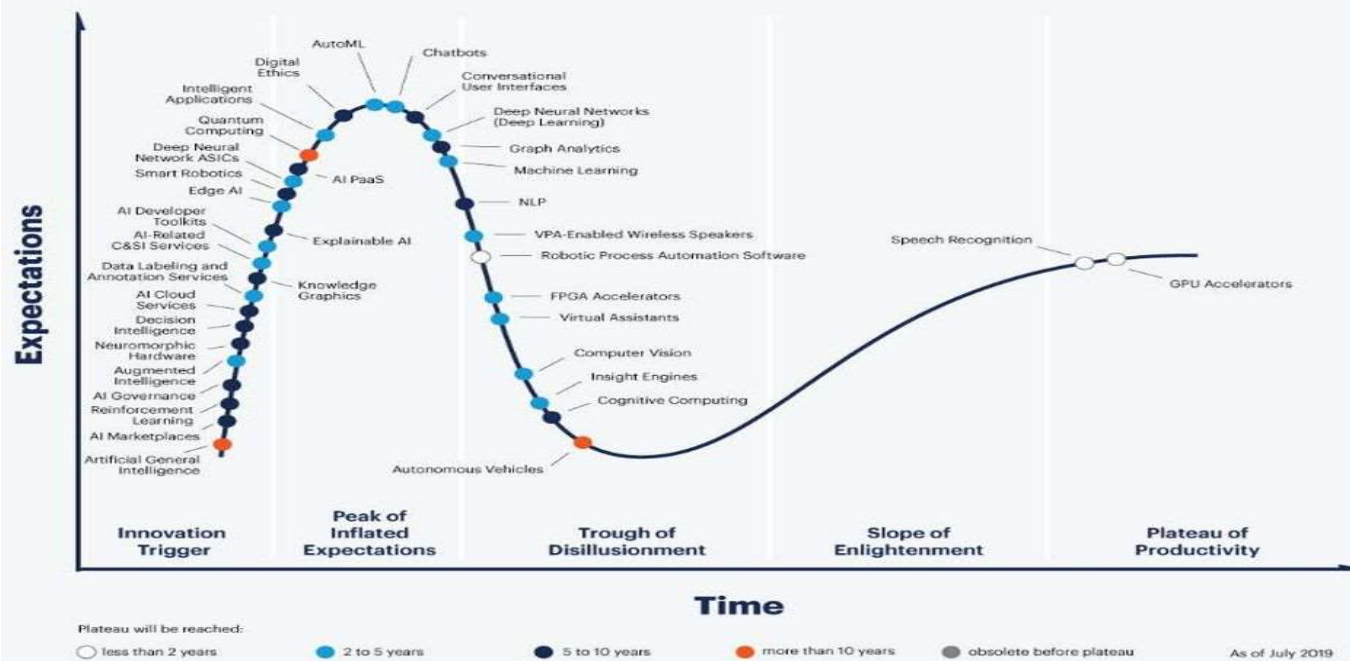
3. 결과



# 1. 주제

- 토픽 모델링을 이용한 인공지능(A.I.) 기술 동향 예측

## Gartner Hype Cycle for Artificial Intelligence, 2019



[gartner.com/SmarterWithGartner](https://gartner.com/SmarterWithGartner)

Source: Gartner  
© 2019 Gartner, Inc. and/or its affiliates. All rights reserved.

**Gartner**

# 1. 주제

---

- 증가하는 인공지능 관련 산업 규모
  - ✓ 글로벌 시장조사기업 옴디아 트랙티카(Omdia Tractica)의 조사 보고서에 따르면 전 세계 AI 엣지 칩셋 시장이 지난해 70억 달러(약 8조7천억원)에서 오는 2025년에는 591억 달러(약 73조2천억원)로 증가할 것이라고 예측함.
- 인공지능 기술 관련한 국면은 이론적, 학문적 연구에서 이제는 산업적 측면으로 전환되고 있다고 함(WIPO Technology Trends 2019, WIPO, 2019).
- 2013년 이후로 빠른 속도로 인공지능 관련 특허들이 출원되고 있고, 학문적 연구들과 특허의 비율이 2010년 8:1의 비율에서, 2016년에는 3:1의 비율로 줄어든 것에서 알 수 있듯이 빠른 속도로 AI 기술이 산업에 적용되고 있다고 함(WIPO Technology Trends 2019, WIPO, 2019).
- 따라서, 인공지능 관련 특허에서 핵심 기술을 파악할 수 있으면 어떤 기술이 유망한 기술인지 알 수 있음.

## 2. 데이터 - 수집

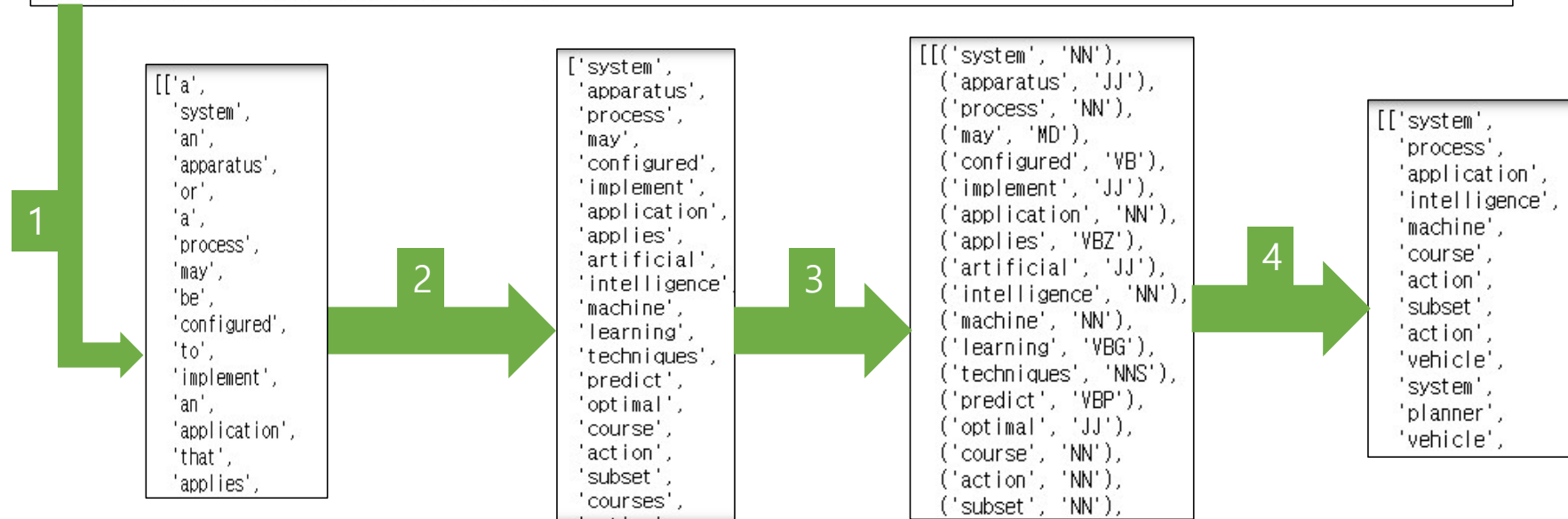
- 미국특허청(USPTO)의 특허 데이터를 이용
- 인공지능(A.I.)과 관련된 키워드를 정리해 관련 특허 수집
  - ✓ 'artificial intelligence', 'deep learning', 'machine learning', 'machine translation', 'robotics', 'NLP', 'neural network' -> 7개의 키워드
- 2010년부터 출원된 총 7830개의 특허 추출
  - ✓ 'artificial intelligence': 689개
  - ✓ 'deep learning': 385개
  - ✓ 'machine learning': 3176개
  - ✓ 'machine translation': 728개
  - ✓ 'robotics': 181개
  - ✓ 'NLP': 244개
  - ✓ 'neural network': 2427개
- 특허 중 util patent만 추출
  - ✓ 총 특허의 개수: 7816개
- 7816개의 데이터에서 patent\_abstract열의 데이터만 추출해서 분석 진행

Unnamed: 0	patent_number	patent_title	patent_date	patent_type	patent_abstract	ipc_list	
0	0	10001775	Machine learning systems and techniques to opt...	2018-06-19	utility	A system, an apparatus or a process may be con...	['G05D', 'G06N']
1	1	10009332	Method and apparatus for remote identity proof...	2018-06-26	utility	Illustrative methods, processes, and software ...	['G06F', 'G06K', 'H04L']
2	2	10009467	Automatic telephone call bridge into a queue	2018-06-26	utility	Providing an artificial intelligence robot con...	['G06Q', 'H04M']
3	3	10013655	Artificial intelligence expert system for anom...	2018-07-03	utility	An artificial intelligence expert system detec...	['G06F', 'G06N']
4	4	10016688	Secondary-game-mode sessions based on primary-...	2018-07-10	utility	Approaches for secondary-game-mode sessions ba...	['A63F', 'G06F', 'G06N']

## 2. 데이터 - 전처리

- tensorflow.keras.preprocessing.text의 text\_to\_word\_sequence 패키지로 tokenization, stop\_words 제거같은 전처리 진행
  - ✓ Stop\_words는 get\_stop\_words('en')과 따로 설정해둔 ['e', 'g', 'etc']를 합쳐서 사용
  - ✓ Pos\_tag 후 "NN", "NNP"를 품사로 가지는 단어(명사)만 추출

```
['A system, an apparatus or a process may be configured to implement an application that applies artificial intelligence and/or machine-learning techniques to predict an optimal course of action (or a subset of courses of action) for an autonomous vehicle system (e.g., one or more of a planner of an autonomous vehicle, a simulator, or a teleoperator) to undertake based on suboptimal autonomous vehicle performance and/or changes in detected sensor data (e.g., new buildings, landmarks, potholes, etc.). The application may determine a subset of trajectories based on a number of decisions and interactions when resolving an anomaly due to an event or condition. The application may use aggregated sensor data from multiple autonomous vehicles to assist in identifying events or conditions that might affect travel (e.g., using semantic scene classification). An optimal subset of trajectories may be formed based on recommendations responsive to semantic changes (e.g., road construction).',
```



### 3. 결과

- Topic 개수를 7개로 지정

	Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7
Top1	model(2225)	device(1830)	memory(969)	language(2121)	network(3346)	image(5107)	unit(893)
Top2	machine(1977)	system(1684)	network(795)	translation(1590)	output(1744)	network(1726)	machine(857)
Top3	system(1309)	machine(1333)	machine(702)	system(1439)	input(1534)	method(932)	state(679)
Top4	plurality(1043)	information(1310)	system(648)	machine(1192)	layer(1043)	feature(840)	device(605)
Top5	training(811)	network(1207)	processor(417)	source(829)	plurality(925)	system(824)	control(577)

- Topic 1: model, machine 등의 단어가 있으므로 학습모델과 관련된 기술일 것
- Topic 2: device, system 등의 단어가 있으므로 ai를 구현하는 장치 및 시스템 관련 기술일 것
- Topic 3: memory, processor 등의 단어가 있으므로 ai 및 머신러닝에 필요한 메모리 관련 기술일 것
- Topic 4: language, translation 등의 단어가 있으므로 ai의 번역 관련 기술일 것
- Topic 5: network, layer 등의 단어가 있으므로 neural network와 관련된 기술일 것
- Topic 6: image의 빈도수가 압도적이므로 딥러닝의 image처리 관련 기술일 것
- Topic 7: unit, control 등의 단어가 있으므로 ai와 관련된 robotics에 관한 기술일 것