

# 텍스트마이닝을 활용한 디지털 정책 주요 의제 분석

이종용



본 보고서는 ETRI 기술정책연구본부 기본사업인  
“국가 지능화 기술정책 및 표준화 연구”를 통해 작성된 결과물입니다.





# 목 차

C O N T E N T S

<b>핵심 요약</b>	<b>1</b>
<b>I. 연구개요</b>	<b>8</b>
1. 연구목적	8
2. 연구내용	8
<b>II. 디지털 정책의 전반적 동향</b>	<b>9</b>
1. 디지털 정책 분야 선정	9
2. 디지털 정책 문서출현 동향	11
3. 인공지능 정책 문서출현 동향	12
<b>III. 디지털 정책 주요 의제 분석</b>	<b>13</b>
1. 연구 방법론	13
2. 분석자료 및 분석절차	16
3. 키워드 분석	17
3. 미래신호 분석	21
4. 토픽모델링	24
5. 디지털 정책 의제와 디지털 국제규범 비교	29
<b>IV. 맺음말</b>	<b>33</b>
1. 결론	33
2. 향후 연구 방향	35
<b>참고문헌</b>	<b>36</b>





## 핵심 요약

 연구개요

## ○ 연구목적

- ‘디지털 정책’은 디지털 전환 관련 정부차원의 선제적인 정책수립 활동으로서 디지털화로 인해 제기되는 기술 및 사회·경제 전반의 다양한 이슈에 대한 국가적 대응 전략을 의미함
- 본 연구는 디지털 전환 시대를 맞이하여 데이터 기반으로 디지털 정책 분야 주요 의제를 도출하여 상호 간 관계구조를 파악하고, 나아가 도출된 디지털 정책 의제와 디지털 경제 국제 규범과의 연계를 확인하는 것을 목적으로 함

## ○ 연구내용

- 디지털 정책 분야 주제군을 정하고 인공지능(artificial intelligence: AI) 정책과 디지털 정책 분야별 문서출현 동향을 검토하여 주요국의 정책적 관심이 집중되는 영역을 파악
- 텍스트마이닝 기법을 활용하여 디지털 정책 관련 핵심 키워드 현황, 미래 신호 관련 키워드를 확인하고, 토픽분석을 통해 디지털 정책 주요 정책 의제를 도출한 다음 이를 최근 타결된 디지털 경제 국제규범과 비교해 상호 연계성을 파악

 디지털 정책의 전반적 동향

- 디지털 정책 도메인(domain)을 제시한 국내·외 자료를 토대로 본 연구는 디지털 정책 분야를 디지털 규제, 디지털 안전, 디지털 기반, 디지털 산업, 디지털 국제화 등으로 구분
- DPA(2022)의 문서출현 빈도를 기준으로 디지털 정책 분야별 글로벌 동향을 살펴본 결과, 데이터 거버넌스 영역을 포함하는 디지털 기반 분야가 가장 큰 비중을 차지하였음
  - ※ DPA(Digital Policy Alert, 2022)는 미국, 중국, EU, 영국, 한국, 일본 등 74개국(OECD, G20, G7, WTO 등 6개 국제기구 포함)을 대상으로 디지털 경제에 영향을 미치는 정부 정책·규제 변화에 관한 법제 및 행정정보를 정리한 데이터베이스이며, Global Trade Alert(GTA)에서 제공하고 있음
- 시간변화에 따른 문서출현 동향의 경우 디지털 규제, 디지털 안전, 디지털 기반 분야가 꾸준히 높은 비중을 보였으며, 무엇보다도 디지털 안전 분야의 정책문서가 '20년부터 '22년 상반기 동안 빠르게 증가한 것으로 나타났음

**표** DPA(2022)의 디지털 정책 문서출현 현황(단위 : 건수)

구분	2018년 이전	2018년	2019년	2020년	2021년	2022년 상반기
디지털 규제 (경쟁, 콘텐츠 규제)	5 (16.1%)	3 (18.8%)	11 (30.6%)	53 (24.3%)	235 (23.8%)	164 (29.9%)
디지털 안전 (사이버보안)	-	1 (6.3%)	3 (8.3%)	8 (3.7%)	82 (8.3%)	63 (11.5%)
디지털 기반 (데이터 거버넌스, 운영환경, 등록 면허, 기타)	23 (74.2%)	5 (31.3%)	14 (38.9%)	106 (48.6%)	532 (54.0%)	281 (51.2%)
디지털 산업 (보조금 산업정책, 공공조달, FDI)	2 (6.5%)	4 (25.0%)	6 (16.7%)	31 (14.2%)	60 (6.1%)	19 (3.5%)
디지털 국제화 (국제통상, 조세)	1 (3.2%)	3 (18.8%)	2 (5.6%)	20 (9.2%)	77 (7.8%)	22 (4.0%)
합계	31 (100%)	16 (100%)	36 (100%)	218 (100%)	986 (100%)	549 (100%)

- 인공지능 정책 문서는 116개이며, 국가별로는 AI 규제에 적극적인 EU(23건)의 정책문서가 가장 많고, 이후 미국(19건), 영국(10건), 중국(8건) 순으로 나타남
- '21년에 많은 문서가 출현한 것은 인공지능 규제 논의 확대와 미국, 중국 등 인공지능 관련 글로벌 기술패권 경쟁의 심화에 따른 것으로 추정

**표** DPA(2022)의 인공지능 정책 문서출현 현황(단위 : 건수)

순번	구분	2018년 이전	2018년	2019년	2020년	2021년	2022년 상반기	합계
1	EU	-	1	-	5	14	3	23
2	미국	-	-	1	2	11	5	19
3	영국	-	-	1	2	6	1	10
4	중국	1	-	-	-	6	1	8
5	호주	-	-	2	1	4	-	7
6	인도	-	-	-	4	1	1	6
7	캐나다	-	-	-	-	4	1	5
8	독일	-	1	-	1	2	-	4
9	한국	-	-	1	1	2	-	4
10	브라질	-	-	-	1	2	-	3
11	기타	1	2	2	3	12	7	27
	합계	2	4	7	20	64	19	116

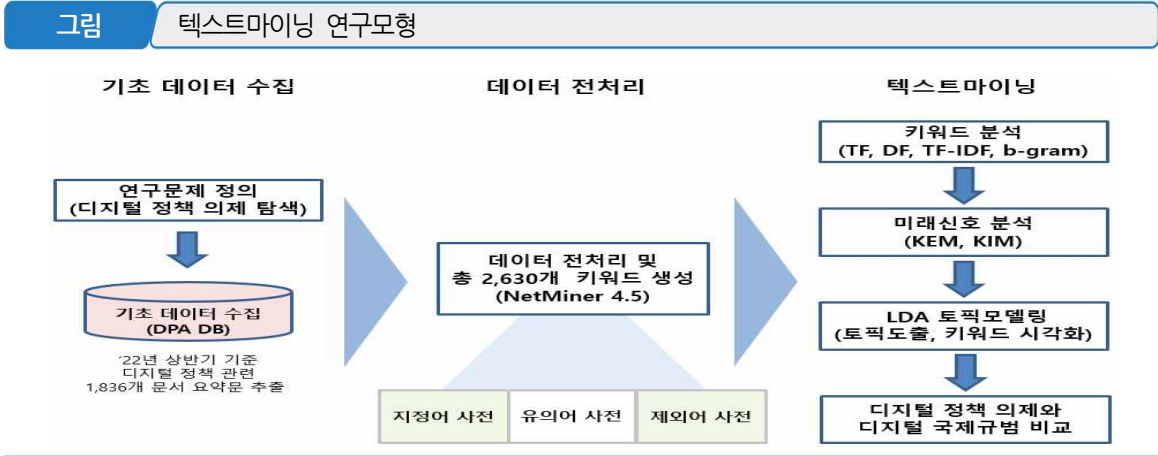
## ④ 디지털 정책 주요 의제 분석

### ○ 연구 방법론

- (분석 데이터) 디지털 정책 채택 동향에서 활용한 '22년 상반기 기준으로 수집한 DPA(2022)에 포함된 1,836개 문서의 요약문을 텍스트마이닝 문서집합으로 추출
- (데이터 전처리) 분석대상 키워드는 소문자로 변환된 영어 명사로 한정하였으며, 제외어, 유의어, 지정어 등 사용자 사전을 통해 기초 데이터를 정제하여 총 2,630개 키워드를 생성
- (분석기간) 연도별 분석에서는 전체기간을 '18년 이전, '18년, '19년, '20년, '21년, '22년 상반기 등 6개 기간으로 구분하였음



- (분석도구) 텍스트마이닝 분석에는 넷마이너(NetMiner) 4.5 프로그램을 사용
- (분석절차) 기초 데이터 수집, 데이터 전처리, 키워드 분석, 미래신호 분석, LDA 기반 토픽 모델링, 토픽별 주요 단어 시각화, 토픽 주제 명명, 디지털 국제규범과 비교 순으로 진행



## 키워드 분석

- data, information, protection, privacy, right 등 가장 많은 정책문서의 출현 비중을 보인 데이터 거버넌스 관련 단어들이 상위를 차지하였으며, 'artificial intelligence' 키워드는 단어 빈도(TF) 기준 27위, 문서빈도(DF) 기준 29위, TF-IDF 가중치에 의한 단어빈도 기준 25위 등 전반적으로 Top 10 순위에는 포함되지 않았음
- 연도별 출현빈도 순위의 경우 '18년 이전에는 internet이 가장 많이 등장하였으나, 이후에는 data가 가장 최상위에 위치하였음
- 인공지능은 '18년(6위)과 '19년(10위)에 비교적 높은 순위에 보였는데, 이것은 해당 시기에 인공지능에 대한 사회적 관심이 고조되었고, 한국, 싱가포르 등 주요국 AI 전략 발표가 영향을 미친 것으로 추정

**표 디지털 정책 문서의 키워드 분석(상위 10개 기준)**

순위	TF		DF		TF-IDF	
	키워드	빈도	키워드	빈도	키워드	빈도
1	data	2060	data	664	data	479
2	information	794	information	468	information	241
3	protection	594	protection	394	platform	234
4	platform	552	security	281	consumer	209
5	security	432	platform	272	protection	197
6	consumer	416	consumer	236	cybersecurity	185
7	cybersecurity	335	privacy	219	security	179
8	privacy	332	right	211	privacy	166
9	content	309	technology	194	competition	161
10	right	307	access	189	content	160

\* 주: TF : Term Frequency, DF : Document Frequency, TF-IDF : Term Frequency-Inverse Document Frequency)

- Bi-gram 분석을 통해 최빈출 키워드와 연관된 키워드의 동시 출현빈도를 살펴본 결과, 'data'의 연관단어 동시 출현빈도는 'data protection'(291회), 'data processing'(85회), 'data security'(81회), 'data transfers'(78회) 순으로 나타났으며, 대체로 데이터 보호·관리 관련 키워드의 동시 출현이 가장 많이 등장하였음

**표** 데이터 관련 Bi-gram 분석

순위	Cross-cutting		유형/형태		수집/저장		가공/처리		이전/공유		보호/관리	
	Bi-gram	N	Bi-gram	N	Bi-gram	N	Bi-gram	N	Bi-gram	N	Bi-gram	N
1	data governance	16	data subject	49	data collection	33	data processing	85	data transfers	78	data protection	291
2	data infrastructure	6	data context	3	data storage	15	data handler	6	data portability	16	data security	81
3	data strategy	3	data information	3	data access	14	data algorithm	4	data sharing	14	data controller	74
4	data platform	2	data russians	3	data cloud	9	data handling	3	data third party	13	data breach	51
5	data artificial intelligence	1	data set	2	data retention	8	data minimization	3	data localization	9	data privacy	46
6	data cooperation	1	data type	2	data individual	5	data analysis	2	data flows	7	data processor	29
7	data economy	1	data activity	1	data consent	4	data communication	2	data consumer	6	data broker	17
8	data partnership	1	data advertising	1	data backup	2	data development	2	data exchange	5	data right	9
9	data stipulates	1	data bulk	1	data generation	2	data activation	1	data exporter	5	data deletion	6

\* 주: Bi-gram은 Source 키워드와 Target 키워드를 공백으로 연결한 복합어로 구성된 것

- 인공지능은 strategy(7회), machine learning(6회), internet of things(4회), 보안(security)은 risk(13회), standard(11회), assessment(9회), 사이버보안은 incident(19회), infrastructure(18회), risk(16회), 플랫폼은 youtube(11회), work(9회), accountability(4회), 경쟁은 consumer(21회), market(19회), concern(4회) 순으로 동시출현 빈도가 높은 것으로 나타났음

**표** AI, 보안, 사이버보안, 플랫폼, 경쟁 관련 Bi-gram 분석

순위	Artificial Intelligence			Security			Cybersecurity			Platform			Competition		
	Source	Target	N	Source	Target	N	Source	Target	N	Source	Target	N	Source	Target	N
1	artificial intelligence	strategy	7	security	risk	13	cybersecurity	incident	19	platform	youtube	11	competition	consumer	21
2		machine learning	6		standard	11		infrastructure	18		work	9		market	19
3		internet of things	4		assessment	9		risk	16		accountability	4		concern	4
4		data	4		threat	8		standard	12		economy	4		digital markets	3
5		governance	3		technology	6		center	10		market	4		meta	3
6		product	3		breach	6		threat	8		worker	4		in app	3
7		industry	2		protection	6		audit	5		news	3		internet	3
8		technology	2		vulnerability	5		data	4		monopoly	3		distortion	2
9		cloud	2		standardisation	4		strategy	4		online marketplace	2		network	2
10		algorithm	2		privacy	4		vulnerability	4		website	2		enforcement	2

### 미래신호 분석

- 강하지는 않지만 잘 알려진 신호에는 market, access, software, social media, payment, investment, google, artificial intelligence 등 주로 디지털 규제, 디지털 기반, 디지털 산업 관련 키워드들이 선정
- 강신호에는 safety, platform, power, competition, child, health, security, infrastructure 등 주로 디지털 규제, 디지털 안전 관련 키워드들이 포함

- 약신호에는 cryptocurrency, supervision, distribution, channel, digital services act, design, party, exchange 등 주로 디지털 기반, 디지털 산업 관련 키워드들이 포함

**표** 디지털 정책 분야 미래신호(상위 20개 키워드 기준)

구분	약신호 (weak signal)	강신호 (strong signal)	강하지는 않지만 잘 알려진 신호 (not strong but well known signal)
KEM	cryptocurrency, supervision, distribution, channel, digital services act, goal, age, challenge, relation, attack, design, party, exchange, site, web, address, category	meta, house, safety, asset, platform, apple, power, consumer, competition, child, minor, health, processing, security, advertising, youtube, individual, infrastructure, crypto asset	transparency, market, assessment, currency, innovation, access, software, cyber, social media, control, part, payment, cooperation, app store, right, investment, google, enforcement, artificial intelligence, application
KIM	meta, supervision, distribution goal, crypto currency, exchange, age, relation, moderation, design, channel, challenge, stakeholder, address, party, web, credit, attack, cyber attack, price	house, platform, asset, power, safety, reason, competition, consumer, processing, individual, health, security, restriction, information, advertising, bank, account, citizen, child, infrastructure	fine, market, medium, google, currency, transfers, transaction, assessment, access, transparency, payment, cyber, part, tax, cooperation, enforcement, software, cloud, artificial intelligence, purpose
미래신호	crypto currency, supervision, distribution, channel, digital services act, goal, age, challenge, relation, attack, design, party, exchange, site, web, address, category	house, safety, asset, platform, power, consumer, competition, child, health, processing, security, advertising, individual, infrastructure	transparency, market, assessment, currency, access, software, cyber, social media, part, payment, cooperation, right, investment, google, enforcement, artificial intelligence, application

- \* 주1): KEM : Keyword Emergency Map, KIM : Keyword Issue Map
- \* 주2): 평균 단어빈도와 가시성(visibility) 정도를 매핑한 KEM과 평균 문서빈도와 확산성(diffusion) 정도를 매핑한 KIM에서 동시에 확인된 약신호, 강신호, 강하지는 않지만 잘 알려진 신호들을 중심으로 미래신호를 도출

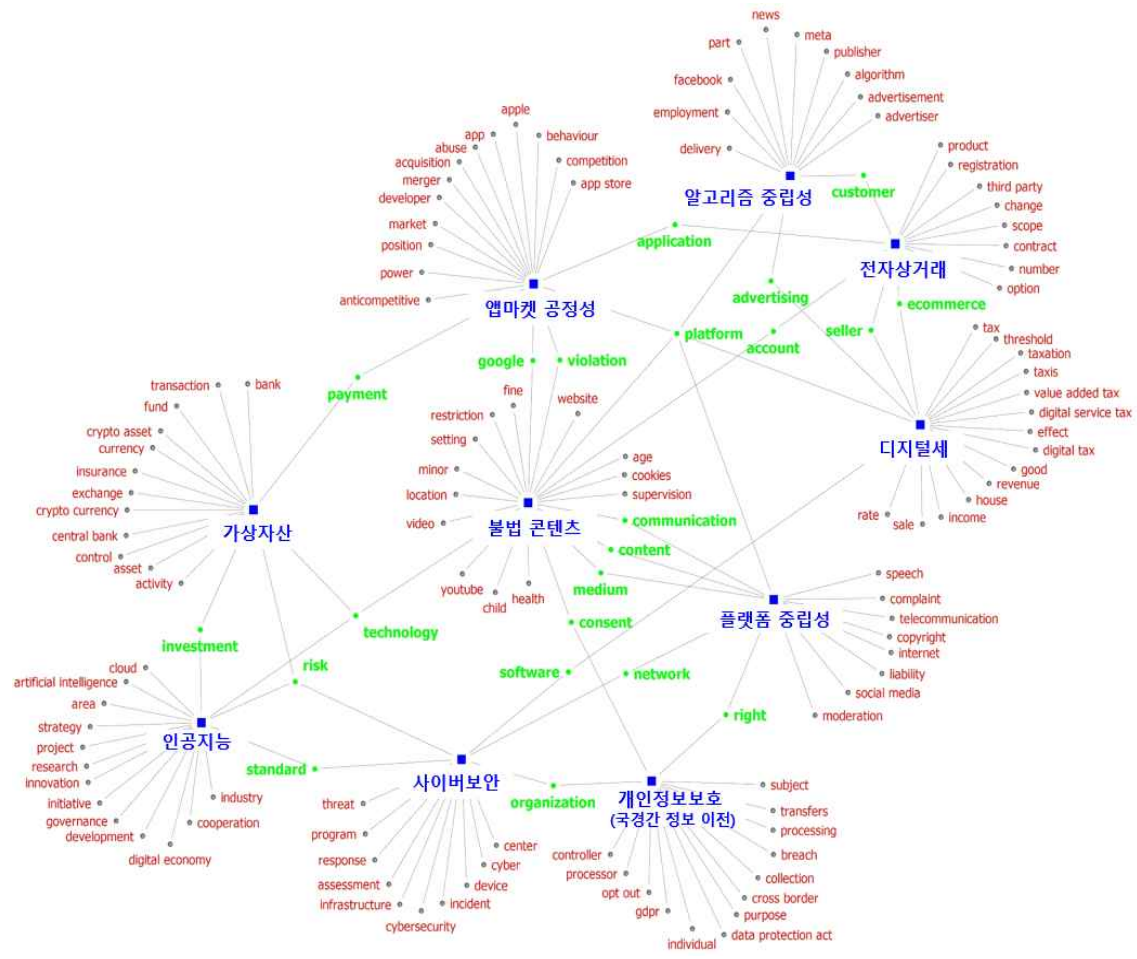
## 토픽모델링

- LDA 토픽분석에 앞서 유의어, 지정어, 제외어 등 사전을 구성하는 작업을 통해 1차 전처리를 마친 단어집합에 대해 TF-IDF 가중치를 토대로 다수 문서에 등장하여 중요도가 낮은 일반적인 키워드를 제거하는 2차 전처리 작업을 수행
- 본 연구는 실루엣 계수(Silhouette coefficient) 평균을 참고해 주제 범주화가 가장 잘 되었다고 판단되는 토픽수(10개)를 선정하였으며, 도출되어진 토픽별 주제는 주요 단어를 참고해 명명하였음
- LDA 토픽분석 결과, 플랫폼 중립성, 인공지능, 앱마켓 공정성, 알고리즘 중립성, 불법 콘텐츠, 사이버보안, 전자상거래, 개인정보보호(국경간 정보 이전), 가상자산, 디지털세 등 10개의 디지털 정책 분야 의제를 도출
  - '21년 기준 토픽비중(문서비중)은 가상자산(17.5%), 앱마켓 공정성(12.3%), 플랫폼 중립성(12.0%), 전자상거래(11.5%), 알고리즘 중립성(9.6%) 순이며, '20년과 '21년 사이에 토픽비중이 증가한 주제는 알고리즘 중립성, 불법 콘텐츠, 사이버보안, 전자상거래, 가상자산 등으로 나타났는데, 이것은 해당 분야의 정책 관심도가 타 분야보다 상대적으로 높아진 것을 보여주고 있음
  - 시각화 네트워크의 토픽 간 관계구조에서 인공지능 토픽은 가상자산-사이버보안과는 'risk', 가상자산-불법 콘텐츠와는 'technology' 키워드를 매개로 연결되며, 가상자산 토픽과는 'investment', 사이버보안 토픽과는 'standard'를 매개로 연결되어 있음

**표** 디지털 정책 분야 주요 토픽(의제) 요약

토픽	주제 요약	주요 단어 (상위 5개)	토픽비중 ('21년)	토픽비중 추세 ('20년~'21년)
1	플랫폼 중립성	content, platform, social media, communication, internet	12.0%	감소
2	인공지능	artificial intelligence, technology, development, strategy, cloud	8.1%	감소
3	앱마켓 공정성	competition, market, google, app store, app	12.3%	-
4	알고리즘 중립성	platform, advertising, facebook, advertisement, algorithm	9.6%	증가
5	불법 콘텐츠	cookies, child, website, communication, technology	7.4%	증가
6	사이버보안	cybersecurity, risk, infrastructure, incident, network	7.0%	증가
7	전자상거래	product, ecommerce, application, number, contract	11.5%	증가
8	개인정보보호 (국경간 정보 이전)	right, processing, transfers, gdpr, controller	5.1%	-
9	가상자산	payment, asset, transaction, bank, investment	17.5%	증가
10	디지털세	tax, revenue, sale, platform, threshold	9.5%	-

**그림** 디지털 정책 분야 토픽별 주요 단어 시각화



## ④ 디지털 정책 의제와 디지털 국제규범

- 한싱 DPA(한국-싱가포르 디지털 동반자 협정, '21년 12월 타결)를 기준으로 디지털 정책 의제와 디지털 분야 국제규범 간 상호 연계성은 확인되나, 다수의 국제규범이 '연성법(soft law)' 단계에 있어 실체법적인 의무가 포함될 때까지는 상당한 시간이 필요

**표** 디지털 정책 의제와 디지털 국제규범 연계

분야	디지털 정책 의제	디지털 경제 국제규범 (한싱 DPA 기준)
디지털 규제	플랫폼 중립성, 앱마켓 공정성, 알고리즘 중립성	전자상거래를 위한 인터넷 접근·이용 원칙, 디지털 경제에서의 경쟁
디지털 안전	불법 콘텐츠 사이버보안 개인정보보호	<b>원치 않는 상업용 전자 메시지(스팸 메시지)</b> 사이버보안 협력, 온라인 안전·보안 <b>개인정보보호</b>
디지털 기반	전자상거래 (전자적 거래 기반조성·신뢰형성)	<b>전자인증·전자서명</b> , 디지털 신원, <b>디지털제품 비차별대우</b> , 디지털 경제를 위한 표준·기술규정·적합성 평가절차, <b>온라인 소비자 보호</b> , <b>소스코드</b> , <b>암호기법을 사용하는 ICT 제품</b> ,
디지털 산업	인공지능 전자상거래 가상자산	인공지능 중소기업·스타트업, 데이터 혁신, 정부 데이터 공개 핀테크 협력
디지털 국제화	디지털세 개인정보보호 (국경간 정보 이전)	<b>관세(전자적 전송 무관세)</b> 전자적 수단에 위한 국경간 정보 이전, 컴퓨팅 설비의 위치

\* 주: 밑줄과 함께 진하게 표현된 조항이 구속력이 있는 강행규범

## ④ 시사점

- 디지털 정책 분야를 5개 주제군(디지털 규제, 디지털 안전, 디지털 기반, 디지털 산업, 디지털 국제화)으로 정하고, G20 회원국을 포함한 주요국의 최신 디지털 정책자료를 토대로 분야별 정책문서의 출현 동향을 분석하였으며, 텍스트마이닝 기법을 활용하여 디지털 정책 관련 주요 키워드별 빈도 현황과 미래신호로 잡히는 키워드를 확인하였음
  - '인공지능'은 사회적인 관심이 고조되었던 '18년과 '19년에 주요 키워드로 등장하여 미래신호 분석에서 '강하지 않지만 잘 알려진 신호'로 탐색이 되었음
- LDA 토픽분석을 통해 도출된 디지털 정책 분야 10대 의제를 디지털 경제 국제규범과 비교해본 결과, 상호 간 유의미한 연계성이 존재하는 것으로 나타났음
  - '인공지능' 토픽은 '21년 기준으로 8.1%의 토픽비중을 차지하였으며, 토픽별 주요 단어 네트워크에서 가상자산과 사이버보안 토픽과 'risk', 가상자산과 불법 콘텐츠 토픽과 'technology' 등을 매개 키워드로 긴밀한 연결관계 구조를 보였음
  - '인공지능' 토픽은 디지털 국제규범 논의에서 '인공지능'이라는 조항으로 명문화는 되었으나, 협정국 상호 간 노력을 촉구하는 단계는 수준이며, 실체법적 의무조항으로 되기까지는 상당한 시간이 필요
- 연계 가능 점점 발굴 등 디지털 전략, AI 전략, 디지털 국제규범 상호 간 일관성을 제고하는 정책적 고려가 디지털 정책의 전략적 효과를 극대화하는 핵심 포인트임을 제언함

# I 연구 개요

## 1 연구목적

- ‘디지털 정책’은 디지털 전환(digital transformation) 관련 정부차원의 선제적인 정책수립 활동으로서 디지털화로 인해 제기되는 기술 및 사회·경제 전반의 다양한 이슈에 대한 국가적 대응 전략을 의미함<sup>1)</sup>
  - 디지털 전환은 1차 산업(농·축·수산업), 제조업, 서비스업 등 전 산업 분야에서 그리고 대·중소기업, 공공 및 개인까지 사회를 구성하는 전 영역에서 이미 상당히 진행되었으며, 향후 더욱 확대될 전망
  - 디지털 전환의 동인은 기술, 데이터, 비즈니스 모델이며, 이를 둘러싼 글로벌 패권 경쟁은 기존의 ‘첨단 산업의 생산기지 국내화’와 ‘프렌드 쇼어링(friend-shoring)’ 형태에서 ‘자국 산업 우선주의’와 ‘보호무역주의’가 가미되어 지속 중임
- 디지털 전환의 기회와 도전에 효과적으로 대응하기 위해서는 디지털 정책 분야의 국제적 움직임에 관한 이해가 필요
  - 주요국은 안보, 경제적 변영, 민주주의 등 디지털 활용의 편익을 극대화하고, 개인정보 남용, 사이버보안 공격, 불공정 행위, 가짜뉴스 등 디지털로 인한 위험을 통제하기 위한 국가 차원의 정책 추진과 동시에 새로운 국제규범 제정을 위해 노력하고 있음
  - 디지털 국제규범 분야에서는 디지털 전환 추세를 반영하여 디지털 기반 국가간 교역을 촉진하는 규범 채택 및 AI 등 디지털 신기술 관련 협력 기반 마련이 논의되고 있음
- 본 연구는 디지털 전환 시대를 맞이하여 데이터 기반으로 디지털 정책 분야 주요 의제를 도출하여 상호 간 관계구조를 파악하고, 나아가 도출된 디지털 정책 의제와 디지털 경제 국제규범과의 정책 공간적 연계를 확인하는 것을 목적으로 함

## 2 연구내용

- 디지털 정책 분야 주제군을 정하고, 미국, EU, 중국, 영국 등 74개국(6개 국제기구 포함)을 대상으로 디지털 정책 및 인공지능 정책 관련 문서출현 동향을 검토하여 국제적으로 정책적 관심이 집중되는 영역을 파악
- 텍스트마이닝 기법을 활용하여 디지털 정책 관련 핵심 키워드의 출현빈도 현황, 미래 신호를 확인하고, 토픽분석을 통해 디지털 정책 주요 의제를 도출함
- 디지털 경제 관련 국제규범과 토픽분석을 통해 도출된 디지털 정책 의제와 비교해 상호 간 연계 가능성을 파악

1) OECD(2022a)는 국가 디지털 전략(national digital strategies)을 디지털 전환에 영향을 주거나 받는 정책 도메인에 걸쳐 디지털 정책 이슈를 해결하는 “국가 수준의 포괄적 디지털 전략”으로 정의하고, 여기에 브로드밴드 개발, AI, IoT, 디지털 안전, 디지털 정부, 인력 등 특정한 섹터, 정책 도메인, 기술에 중점을 두는 전략정책을 포함

## II 디지털 정책의 전반적 동향

### 1 디지털 정책 분야 선정

- 디지털은 경제의 전 부문을 변화시켜 엄청난 경제적 가치와 전례가 없는 기회를 창출하는 긍정적 측면과 함께 새로운 위험과 도전을 초래하고 있음
  - 디지털이 가져올 광범위한 기회와 도전의 양면성에 대응하여 주요국에서 채택·시행되고 있는 디지털 정책 분야별 동향을 파악할 필요가 있음
- 디지털 정책 도메인 관련 국내·외 자료를 토대로 디지털 정책을 5개 분야로 구분
  - ①디지털 규제(ex, 공정경쟁, 콘텐츠 규제, 온라인 광고 규제, 알고리즘 규제), ②디지털 안전(ex, 사이버보안, 통신망 재난), ③디지털 기반(ex, 데이터 거버넌스, 통신 인프라), ④디지털 산업(ex, 산업정책, AI 전략, 공공조달), ⑤디지털 국제화(ex, 디지털세, 디지털 통상, 디지털 연대)

표 1 디지털 정책 이슈 탐색 관련 기존 연구 사례

구분	내용	비고
PwC(2021) DCMS(2022)	<ul style="list-style-type: none"> <li>● 디지털의 고유 특성을 중심으로 7대 주제군(thematic clusters) 제안                             <ul style="list-style-type: none"> <li>● (데이터 거버넌스) 개인정보의 축적, 가공, 이동</li> <li>● (디지털 콘텐츠 이슈) 디지털 콘텐츠의 감독, 책임, 검증</li> <li>● (알고리즘 규제) 데이터 분석·알고리즘의 투명성과 이용</li> <li>● (디지털 경쟁이슈) 디지털 비즈니스의 규모·범위 경제, 네트워크 효과</li> <li>● (디지털 국제화) 데이터 및 디지털의 초국경적 특성</li> <li>● (디지털 기반) 디지털 인프라와 네트워크의 주요한 역할</li> <li>● (디지털 안전) 디지털 서비스 붕괴의 상대적 용이성과 익명성</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>● 디지털 차원을 제품·서비스, 섹터, 직업, 기술, 비즈니스 모델, 활동 등으로 구분</li> <li>● 디지털 차원별 영향요인을 식별하고 정책적 고려가 요구되는 주제군 파악</li> </ul>
유재홍(2022)	<ul style="list-style-type: none"> <li>● 인공지능, 블록체인, 메타버스 등 디지털 신기술 등장과 잠재 위험의 다양화·다변화에 대응한 <b>디지털 안전</b> 대책 확보 필요성 제기</li> </ul>	<ul style="list-style-type: none"> <li>● 기술발전에 따른 디지털 안전 중요성 제시</li> </ul>
Dig.watch Kurbalija(2016)	<ul style="list-style-type: none"> <li>● 인터넷 거버넌스 관련 40~50개 이슈를 <b>기반, 사이버보안, 인권, 법·규제, 경제, 개발, 사회문화</b> 등 7개 바스켓으로 분류</li> <li>● (디지털 정책 이슈) 경쟁규제, 콘텐츠 정책, 지식재산권, 중개사업자의 책임, 중요 인프라, 네트워크 안전, 통신 인프라, 데이터 거버넌스, 표현의 자유, 디지털 비즈니스 모델, 전자상거래 및 통상, 조세 등</li> </ul>	<ul style="list-style-type: none"> <li>● 바스켓별 정책 이슈 제시</li> </ul>
OECD(2020)	<ul style="list-style-type: none"> <li>● <b>액세스, 이용, 혁신, 일자리, 사회, 신뢰, 시장 개방성</b> 등 상호 연결된 7개 디지털 정책 차원을 정의하고, 각 정책 공간(policy space)별로 고려할 필요가 있는 24개 디지털 정책 도메인 제시</li> </ul>	<ul style="list-style-type: none"> <li>● 디지털 정책 차원별 연계 정책 도메인 파악</li> </ul>
DPA(2022)	<ul style="list-style-type: none"> <li>● (디지털 정책 영역) <b>경쟁, 콘텐츠 규제, 데이터 거버넌스, FDI, 국제통상, 운영환경, 공공조달, 등록 및 허가, 보조금·산업정책, 조세, 기타</b></li> <li>● (정책 수단) 알고리즘 디자인 및 기술규정, 운영환경, 플랫폼 공정경쟁, 콘텐츠 규제, 데이터 현지화, 데이터 이전, 데이터 보호 규제, 국경조치, 디지털세, 사이버보안 규제, 금융적 지원, 공공조달 현지화 등</li> </ul>	<ul style="list-style-type: none"> <li>● 보조금·산업정책, 국제통상, 조세 등 디지털 정책영역별 세부 정책 수단 파악</li> </ul>

\* 출처: 기존 연구 사례 기반 저자 작성

- 디지털 정책 분야의 주제군으로 분류된 세부 정책 이슈는 독립적이기보다는 여러 분야와 연계되는 특성이 있는데, 예를 들어, 데이터 거버넌스 이슈는 디지털 규제와 디지털 기반, 디지털 국제화 분야에도 영향을 미치는 공통 이슈에 해당

표 2 디지털 정책 분야 주제군과 관련 정책 이슈

구분	OECD(2020)	Dig.watch	DPA(2022)
디지털 규제	<ul style="list-style-type: none"> <li>경쟁</li> <li>디지털 격차 해소를 위한 사회정책</li> </ul>	<ul style="list-style-type: none"> <li>경쟁규제</li> <li>양종립성 및 제로레이팅</li> <li>지식 재산권</li> <li>중개사업자의 책임</li> <li>컨버전스 및 OTT</li> <li>디지털 레거시(digital legacies)</li> <li>콘텐츠 정책</li> <li>문화적 다양성</li> <li>다중언어(multilingualism)</li> <li>온라인 교육</li> <li>다학제적(Interdisciplinary) 접근</li> <li>온라인 어린이 안전</li> <li>잊혀질 권리</li> <li>표현의 자유</li> <li>장애인 권리</li> <li>온라인 양성평등</li> <li>인권 원칙</li> <li>폭력적 극단주의</li> </ul>	<ul style="list-style-type: none"> <li>알고리즘 디자인 및 기술규정</li> <li>반경쟁적인 협정 규제</li> <li>경쟁 규제 관련 거버넌스</li> <li>링크 지불 조건(link payment requirement)</li> <li>합병규제</li> <li>감독 메커니즘 구조</li> <li>규제 대상자 정의 및 기준</li> <li>미성년자(minors) 관련 규제</li> <li>일방적 행위(unilateral conduct)</li> <li>콘텐츠 필터링 또는 모니터링 규제</li> <li>콘텐츠 규제 관련 거버넌스</li> <li>지식 재산권(저작권 제외)</li> <li>현시 콘텐츠 의무 관련 투자</li> <li>지식 재산권 관련 콘텐츠 책임</li> <li>현시 콘텐츠 요건</li> <li>지식 재산권 이외 관련 콘텐츠 의무</li> <li>소액 지불 요건(snippet payment requirement)</li> <li>이용자 발언권(user speech rights)</li> <li>영업비밀 공개 요건</li> <li>불법 콘텐츠 삭제 규제(take-down regulation)</li> </ul>
디지털 안전	<ul style="list-style-type: none"> <li>디지털 위험 관리</li> <li>디지털 안전</li> </ul>	<ul style="list-style-type: none"> <li>사이버 범죄</li> <li>중요 인프라</li> <li>사이버 갈등 및 전쟁</li> <li>암호화</li> <li>네트워크 안전(network security)</li> </ul>	<ul style="list-style-type: none"> <li>사이버보안 규제</li> </ul>
디지털 기반	<ul style="list-style-type: none"> <li>통신 인프라 및 서비스</li> <li>소비자 보호</li> <li>디지털 정부</li> <li>노동시장 재편</li> <li>고용 약자를 위한 사회적 보호</li> <li>프라이버시</li> <li>인력(skills)</li> <li>세금 및 보장 정책</li> </ul>	<ul style="list-style-type: none"> <li>통신 인프라</li> <li>중요 인터넷 자원</li> <li>디지털 표준</li> <li>프라이버시 및 데이터 보호</li> <li>디지털 신원</li> <li>관할권(jurisdiction)</li> <li>분쟁해결(dispute resolution)</li> <li>데이터 거버넌스</li> <li>소비자 보호</li> <li>액세스</li> <li>포용적 자금지원</li> <li>e-폐기물(e-waste)</li> </ul>	<ul style="list-style-type: none"> <li>데이터 거버넌스 책임</li> <li>데이터 이동성 규제</li> <li>데이터 보호 규제 관련 거버넌스</li> <li>데이터 보호 규제</li> <li>디지털 신원 요건</li> <li>상호운용성 규제</li> <li>현지 사업 요건(local operation requirements)</li> <li>의무적 데이터 공유 요건</li> <li>서비스 품질 요건</li> <li>상품 또는 서비스 허가</li> <li>고용 규제</li> <li>사업 허가</li> <li>의무 SW 설치 요건</li> </ul>
디지털 산업	<ul style="list-style-type: none"> <li>기업 역동성(dynamism)</li> <li>기업가정신(entrepreneurship)</li> <li>환경</li> <li>헬스케어</li> <li>금융시장</li> <li>투자</li> <li>지역개발</li> <li>과학기술</li> <li>산업별 정책</li> <li>중소기업</li> </ul>	<ul style="list-style-type: none"> <li>디지털 비즈니스 모델</li> <li>역량 개발(capacity development)</li> <li>지속가능한 개발</li> </ul>	<ul style="list-style-type: none"> <li>다국적 기업: 금융 인센티브</li> <li>다국적 기업: 대우 및 사업운영</li> <li>다국적 기업: 진입 및 소유 규정</li> <li>공공조달 접근</li> <li>공공조달 블랙리스트(blacklisting)</li> <li>공공조달 현지화(localisation)</li> <li>공공조달 우대마진(preferance margin)</li> <li>자본 투입 및 주식지분(구제금융 포함)</li> <li>재경지원(financial grant)</li> <li>정부 지원금 규제(state aid control)</li> <li>세금 또는 사회적 보험 경감</li> </ul>
디지털 국제화	<ul style="list-style-type: none"> <li>조세</li> <li>통상(trade)</li> </ul>	<ul style="list-style-type: none"> <li>전자상거래 및 통상</li> </ul>	<ul style="list-style-type: none"> <li>수출 허가 요건</li> <li>수입 허가 요건</li> <li>수입품에 대한 내국세(internal taxation of imports)</li> <li>직접세(디지털서비스세 포함)</li> <li>간접세</li> <li>고정사업장 포함 세원 정의(tax base definition)</li> <li>디지털 서비스 관련 세금</li> <li>데이터 저장 의무(data residency obligations)</li> <li>데이터 이전 규제</li> </ul>

\* 주: 기존 연구 사례에서 제시된 정책 이슈를 본 연구의 디지털 정책 분야 주제군에 맞게 재분류한 것이며, 판단기준이나 관점에 따라 각 정책 이슈의 배치는 달라질 수 있음



## 2 디지털 정책 분야별 문서출현 동향

- DPA(2022)를 활용해 디지털 정책 분야별 문서출현 빈도를 살펴본 결과, 데이터 거버넌스 영역을 포함하고 있는 디지털 기반 분야가 가장 큰 비중을 차지
  - ※ DPA(Digital Policy Alert, 2022)<sup>2)</sup>는 Global Trade Alert(GTA)<sup>3)</sup>에서 제공하는 '22년 상반기 기준 1,836개 문서로 구성된 데이터베이스로서, 미국, 중국, EU, 영국, 한국, 일본 등 74개국(OECD, G20, G7, WTO 등 6개 국제기구 포함)의 디지털 경제 관련 정책·규제 변화를 모니터링한 것임
- 시간변화에 따른 문서출현 동향을 살펴보면, 디지털 규제, 디지털 안전, 디지털 기반 분야가 꾸준히 높은 비중을 보였으며, 디지털 안전 분야는 '20년 이후부터 빠르게 증가한 것으로 나타났음
  - 하위 영역에서는 데이터 거버넌스의 문서출현이 가장 많으며, 경쟁, 콘텐츠 규제, 사이버 보안, 데이터 거버넌스, 운영환경, 조세 영역의 문서출현 비중이 증가하는 추세임

표 3 DPA(2022)의 디지털 정책 문서 출현 현황(단위 : 건수)

구분	2018년 이전	2018년	2019년	2020년	2021년	2022년 상반기
디지털 규제	5 (16.1%)	3 (18.8%)	11 (30.6%)	53 (24.3%)	235 (23.8%)	164 (29.9%)
경쟁	2 (6.5%)	1 (6.3%)	4 (11.1%)	18 (8.3%)	122 (12.4%)	70 (12.8%)
콘텐츠 규제	3 (9.7%)	2 (12.5%)	7 (19.4%)	35 (16.1%)	113 (11.5%)	94 (17.1%)
디지털 안전	-	1 (6.3%)	3 (8.3%)	8 (3.7%)	82 (8.3%)	63 (11.5%)
사이버보안	-	1 6.3%	3 8.3%	8 3.7%	82 8.3%	63 11.5%
디지털 기반	23 (74.2%)	5 (31.3%)	14 (38.9%)	106 (48.6%)	532 (54.0%)	281 (51.2%)
데이터 거버넌스	7 (22.6%)	3 (18.8%)	7 (19.4%)	60 (27.5%)	329 (33.4%)	170 (31.0%)
운영환경	14 (45.2%)	2 (12.5%)	5 (13.9%)	35 (16.1%)	139 (14.1%)	89 (16.2%)
등록·면허	2 (6.5%)	-	-	11 5.0%	53 5.4%	20 3.6%
기타	-	-	-	-	11 1.1%	2 0.4%
디지털 산업	2 (6.5%)	4 (25.0%)	6 (16.7%)	31 (14.2%)	60 (6.1%)	19 (3.5%)
공공조달	-	-	-	2 0.9%	6 0.6%	5 0.9%
보조금 산업정책	2 (6.5%)	4 (25.0%)	6 (16.7%)	14 (6.4%)	40 (4.1%)	11 (2.0%)
외국인 직접투자	-	-	-	15 6.9%	14 1.4%	3 0.5%
디지털 국제화	1 (3.2%)	3 (18.8%)	2 (5.6%)	20 (9.2%)	77 (7.8%)	22 (4.0%)
국제통상	-	-	-	-	9 (0.9%)	7 (1.3%)
조세	1 (3.2%)	3 (18.8%)	2 (5.6%)	20 (9.2%)	68 (6.9%)	15 (2.7%)
합계	31 (100%)	16 (100%)	36 (100%)	218 (100%)	986 (100%)	549 (100%)

\* 주: 문서출현 빈도가 낮은 '09년~'17년 기간은 '18년 이전으로 통합하였음

2) <https://digitalpolicyalert.org>  
 3) <https://www.globaltradealert.org>

### 3 인공지능 정책 문서출현 동향

- 본 연구에서 활용한 DPA(2022)에서는 인공지능과 관련 116개 문서가 확인되며, 국가별로는 AI 규제에 대해 적극적인 EU(23건)에서 나온 정책문서가 가장 많고, 이후 미국(19건), 영국(10건), 중국(8건) 순으로 나타남
- '21년에 많은 문서가 출현한 것은 인공지능 규제 논의 확대와 미국, 중국 등 인공지능 관련 글로벌 기술패권 경쟁의 심화에 따른 것으로 추정

표 4 DPA(2022)의 인공지능 정책 문서 출현 현황(단위 : 건수)

순번	구분	2018년 이전	2018년	2019년	2020년	2021년	2022년 상반기	합계
1	EU	-	1	-	5	14	3	23
2	미국	-	-	1	2	11	5	19
3	영국	-	-	1	2	6	1	10
4	중국	1	-	-	-	6	1	8
5	호주	-	-	2	1	4	-	7
6	인도	-	-	-	4	1	1	6
7	캐나다	-	-	-	-	4	1	5
8	독일	-	1	-	1	2	-	4
9	한국	-	-	1	1	2	-	4
10	브라질	-	-	-	1	2	-	3
11	기타	1	2	2	3	12	7	27
합계		2	4	7	20	64	19	116

- AI의 사회적 편익을 극대화하고 AI 기술개발 지원을 목적으로 하는 국가의 정책 문서인 AI 전략의 경우 전체 18개국, G20 기준 16개국의 자료를 포함

표 5 DPA(2022)의 국가별 인공지능 전략 문서 현황

국가	인공지능 전략 현황	발표일
일본	Japan AI strategy: Artificial Intelligence Technology Strategy	2017-03-30
중국	China AI strategy: A Next Generation Artificial Intelligence Development Plan	2017-07-20
멕시코	Mexico Artificial Intelligence Agenda	2018-03-27
프랑스	AI for Humanity: French Strategy for Artificial Intelligence	2018-03-29
독일	German AI strategy: AI Made in Germany	2018-11-01
EU	EU AI strategy: Coordinated Plan on Artificial Intelligence (COM(2018) 795 final)	2018-12-07
미국	Executive Order 13589 Maintaining American Leadership in Artificial Intelligence	2019-02-11
	National Artificial Intelligence Initiative Act of 2020	2021-01-01
영국	UK AI Sector Deal	2019-05-21
	UK AI Strategy	2021-09-22
러시아	Russia AI Strategy in Decree 490	2019-10-10
싱가포르	Singapore National Artificial Intelligence Strategy	2019-11-13
한국	South Korea National Strategy for Artificial Intelligence	2019-12-17
인도	India National Strategy On Artificial Intelligence	2020-06-04
이탈리아	Proposal for an Italian Strategy for Artificial Intelligence	2020-07-02
인도네시아	Indonesia National AI strategy	2020-08-10
스위스	Swiss AI Strategy (Digitale Schweiz)	2020-09-11
독일	Updated German AI strategy	2020-12-02
브라질	Brazilian Artificial Intelligence Strategy	2021-04-06
튀르키예	National Artificial Intelligence Strategy	2021-08-24

\* 주1): Stanford Univ.(2021)에 의하면, '20년 기준으로 32개국이 AI 전략을 발표하였으며, 22개국이 개발 중임  
 \* 주2): AI 관련 정책 플랫폼인 OECD.AI는 69개 국가의 AI 정책 및 전략에 관한 정보를 제공 중임  
 \* 주3): 국가 리스트는 발표일 순이며, G20 국가는 음영으로 표시

# Ⅲ 디지털 정책 주요 의제 분석

## 1 연구 방법론

### ◎ 텍스트마이닝 개요

- 디지털화와 함께 비정형데이터의 가치가 높아지면서 트렌드, 인식변화, 미래예측, 의제 도출 등에 텍스트마이닝 기법이 다양한 학문 분야에서 널리 활용되고 있음<sup>4)</sup>
  - 텍스트 데이터에 잠재된 패턴과 규칙을 탐색하여 유용한 정보를 추출하는 텍스트마이닝은 기계학습 알고리즘을 활용해 시간·비용 측면에서 효율적이며, 질적자료를 양적자료로 변환하므로 비교적 객관적인 결과도출이 가능<sup>5)</sup>

표 6 ICT 분야 텍스트마이닝을 적용한 연구

구분	분석자료	연구목적
박세훈 외(2022)	뉴스, 논문, 연구보고서	디지털 정보격차와 관련된 주요 토픽 탐색
박대영 외(2021)	국내의 학술지에 수록된 논문정보	공공데이터에 관한 연구동향 파악
심동녘 외(2020)	ITFIND 포털 내 분석 보고서	ICT 분야 이슈 탐색
노설현(2020)	뉴스기사	인공지능 관련 이슈 탐색
남현동 외(2020)	국내 공공기관 연구보고서(10개)	플랫폼 정부 관련 연구 동향 모색
우창우 외(2020)	국가연구개발사업 과제정보	ICT 분야 주요 연구 토픽 도출
조혜인 외(2019)	국내외 논문초록	미국, 중국, 한국의 블록체인 관련 연구 동향 비교
서호준(2019)	'과학기술 50년사'의 텍스트 정보	국내 과학기술 관련 주요 의제 탐색
조경원 외(2019)	논문 정보	4차 산업과 관련 연구주제 탐색
서병조 외(2017)	국내 학술논문 정보	플랫폼 정부 관련 연구 동향 탐색

\* 주: 기존 연구 사례 기반 저자 작성

### ◎ 키워드 분석

- 텍스트마이닝에서는 분석대상인 문서집합(corpus)의 단어 출현빈도를 측정하는 단어빈도(Term Frequency: TF)와 특정 단어가 출현하는 문서의 수를 측정하는 문서빈도(Document Frequency: DF) 등 키워드 분석을 기본으로 수행함
- 자주 출현하더라도 의미상 중요성을 부여하기 어려운 경우 단어빈도 TF와 문서집합 내 동시 출현빈도를 고려하는 역문서 빈도(Inverse Document Frequency: IDF)를 결합해 TF-IDF를 산출함으로써 문서별 특정 단어의 중요도를 나타낸 가중치로 활용<sup>6)7)</sup>

4) 행정, 국제외교, 보건환경, 고용 등 경제·사회 분야를 대상으로 이슈, 담론, 특정 주제 파악을 목적으로 추진된 다양한 국내 텍스트 자료 기반 연구 사례에 대해서는 심동녘 외(2020)를 참고할 것  
 5) 김길래 외 (2021), Covid-19 대유행 과정에서 나타난 중소기업 이슈 분석에 관한 연구, 중소기업금융연구, Vol. 41, No. 1, pp. 49-77.  
 6) Cyram (2022), NetMiner 소개와 기초 사용법.  
 7) Christopher et al.(2008), Introduction to Information Retrieval, Chapter 6, Cambridge University Press.

※  $TF-IDF_{t,d} = TF_{t,d} \times \log\left(\frac{N}{DF_t}\right)$ ,  $TF_{t,d}$  = 문서 d에서의 단어 t 빈도,  $DF_t$  = 단어 t의 문서빈도, N = 총 문서 수

※ TF-IDF는 특정 문서에는 높은 빈도를 가지면서 소수의 문서에 출현하는 단어들에 중요성을 부여

- N개의 단어가 연쇄적으로 출현하는 빈도를 측정하는 ‘N-gram 분석’을 통해 단어와 단어의 동시출현 관계를 파악하며<sup>8)</sup>, 통상적으로 ‘Bi-gram’을 활용<sup>9)</sup>
- 키워드 분석은 상대적으로 출현빈도가 높은 주요 단어를 확인하고, 기간별 추세를 통해 트렌드, 인식변화 파악에 활용

### ◎ 미래신호 분석

- 단어빈도와 문서빈도의 시점별 비중을 활용하여 키워드 포트폴리오 지도(keyword portfolio maps)를 작성하고 미래신호를 탐지함<sup>10)</sup>

- 키워드 포트폴리오 지도는 가시성 정도(Degree of Visibility: DoV)를 활용하는 키워드 출현 지도(Keyword Emergency Map: KEM)와 확산 정도(Degree of Diffusion: DoD)를 활용하는 키워드 이슈 지도(Keyword Issue Map: KIM)로 구분

$$\text{※ } DoV_{ij} = \left(\frac{TF_{ij}}{NN_j}\right) \times \{1 - tw \times (n - j)\}, \text{ } DoD_{ij} = \left(\frac{DF_{ij}}{NN_j}\right) \times \{1 - tw \times (n - j)\}$$

※ tw = 최근 출현한 키워드에 더 큰 가중치를 부여하는 시간 가중치, n = 전체 시간구간, j = 특정 시점,  $TF_{ij}$  = j 시점에 단어의 단어빈도,  $DF_{ij}$  = j 시점에 단어의 문서빈도,  $NN_j$  = j 시점에 총 문서수

- KEM과 KIM에서 동시에 확인된 강신호, 약신호, 강하지 않지만 잘 알려진 신호 영역의 키워드를 미래신호로 활용<sup>11)</sup>

- 4분면 평면에서 KEM은 x축에는 평균 단어빈도, y축에는 DoV의 평균증가율(기하평균), KIM은 x축에는 평균 문서빈도, y축에는 DoD의 평균증가율을 나타냄

※ (1사분면) 평균 단어빈도 또는 평균 문서빈도와 DoV의 평균증가율 또는 DoD의 평균증가율이 모두 높은 영역의 키워드를 강신호(strong signal)로 식별

※ (2사분면) 평균 단어빈도 또는 평균 문서빈도가 낮으면서 DoV의 평균증가율 또는 DoD의 평균증가율이 높은 영역의 키워드를 약신호(weak signal)로 식별

※ (3사분면) 평균 단어빈도 또는 평균 문서빈도와 DoV의 평균증가율 또는 DoD의 평균증가율이 모두 낮은 영역의 키워드를 잠재신호(latent signal)로 식별

※ (4사분면) 평균 단어빈도 또는 평균 문서빈도는 높으면서 DoV의 평균증가율 또는 DoD의 평균증가율은 낮은 영역의 키워드를 강하지 않지만 잘 알려진 신호(not strong but well known signal)로 식별

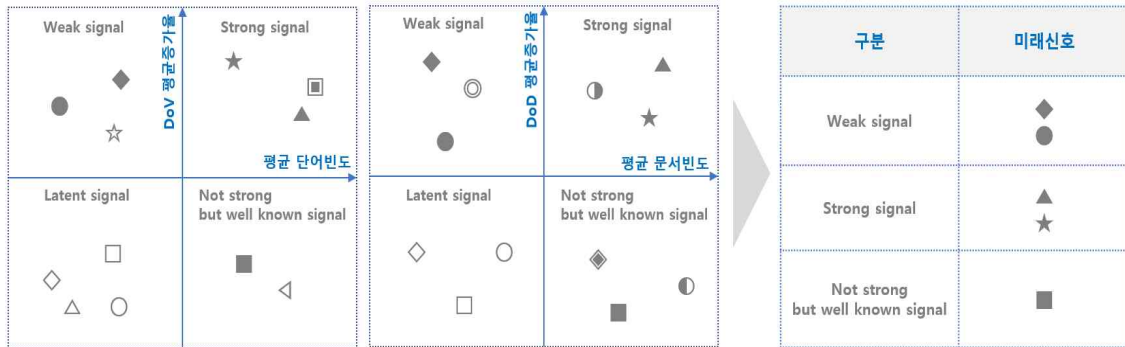
8) ‘N-gram’은 N개의 연속적인 단어 나열 또는 단어 뭉치를 의미하며, N=2일 때가 ‘Bi-gram’임

9) 남현동 외 (2020), 한국 플랫폼 정부의 방향성 모색 : 공공기관 연구보고서에 대한 토픽 모델링과 네트워크 분석, 디지털융복합연구, Vol. 18, No. 2, pp. 139-149.

10) Yoon (2012), “Detecting weak signals for long-term business opportunities using text mining of Web news,” Expert Systems with Applications, Vol. 39(2012), pp. 12543-12550.

11) 박찬국 외 (2015), 사물인터넷을 통한 에너지산업 발전방향 연구: 텍스트마이닝을 이용한 미래 신호 탐색, 기본 연구보고서, 15-21, 에너지경제연구원.

**그림 1** 평면에서 미래신호 식별



\* 출처: Yoon(2012) 기반 저자 작성

- 미래신호 분석은 정의된 키워드 포트폴리오 지도를 통해 키워드 집합의 식별은 매우 직관적이지만, 키워드 집합에 내재된 맥락에 대한 해석은 연구자의 판단에 의존하므로 객관적인 지식을 발견에는 한계가 있음<sup>12)</sup>

### ◎ 토픽모델링

- 비정형데이터에 잠재된 변수(토픽)를 탐색하는 확률모델로서 잠재 디리클레 할당(Latent Dirichlet Allocation: LDA) 기법이 가장 널리 사용되고 있음
  - LDA를 활용할 경우 다량의 키워드들을 동일한 관련 토픽단위로 군집화함으로써 문서 집합에 내재된 복수의 토픽을 모두 고려할 수 있음<sup>13)</sup>
    - ※ m개의 단어수와 n개의 문서수로 구성되어 있는 문서집합이 주어졌을 때 mxn의 단어-문서 행렬을 생성
    - ※ 단어-문서 행렬은 LDA의 투입물이 되어 데이터에 숨겨진 토픽을 확인해 kxn 토픽-문서 행렬을 생성
- 토픽수(k), 샘플링 회수(n), 문서당 토픽 분포계수( $\alpha$ ) 토픽별 단어 분포계수( $\beta$ ) 등이 LDA 결과값에 영향을 미치는 중요한 파라미터임<sup>14)</sup>
  - k,  $\alpha$ ,  $\beta$ 는 LDA 실행 파라미터이며, n은 깁스(Gibbs) 샘플링<sup>15)</sup> 생성모델 파라미터임
  - ※  $\alpha$ 값이 클수록 문서별 토픽이 보다 일정하게 분포되며,  $\beta$ 값 클수록 토픽별 단어가 보다 균일하게 분포
  - ※ 토픽수가 너무 많으면 무의미한 토픽도출이 발생하거나 토픽수가 너무 작으면 한 토픽에 많은 키워드가 몰려 토픽주제 선정이 곤란할 수 있으므로 실무에서는 혼잡도(perplexity), 주제 일관성 점수(coherence score), 실루엣 계수(Silhouette coefficient) 등이 적정 토픽수 k 탐색에 활용되고 있음

12) Krigsholm et al. (2019), Applying text mining for identifying future signals of land administration, Land, Vol. 8, No. 181(Nov. 2019), pp. 1-15.

13) 김성목 외 (2020), 텍스트마이닝을 이용한 리빙랩 연구동향 분석, 디지털융복합연구, Vol. 18, No. 8, pp. 37-48.

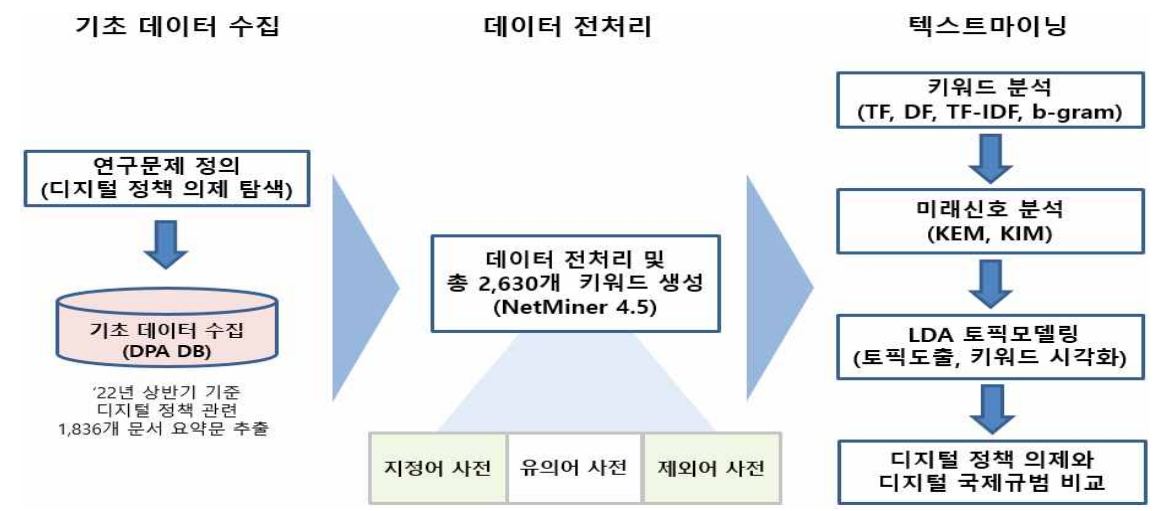
14) Panichella et al. (2013), How to effectively use topic models for software engineering tasks, ICSE 2013, pp. 522-531.

15) 깁스 샘플링은 복수의 확률변수를 갖는 결합확률분포로부터 연속적인 표본을 생성하는 알고리즘임

## 2 분석자료 및 분석절차

- (기초 데이터) 디지털 정책 채택동향에서 활용한 '22년 상반기 기준으로 수집한 DPA(2022)에 포함된 1,836개 문서의 요약문을 텍스트마이닝 문서집합으로 추출
- (데이터 전처리) 분석대상 키워드는 소문자로 변환된 영어 명사로 한정하였으며, 제외어(exception lists), 유의어(thesaurus), 지정어(defined words) 사전을 통해 기초 데이터를 정제하여 총 2,630개 키워드를 생성
  - 지역명, 기관명, 직책, 국가명과 같은 디지털 정책과 관련성이 낮은 키워드, 기타 불용어(stopwords)를 모아 1,285개의 제외어 사전 작성
  - 여러 형태로 조합된 복합명사 그리고 이름, 상호와 같은 고유명사가 정확히 추출되도록 188개의 지정어 사전 작성
  - 유사한 의미의 단어가 대표단어로 추출되도록 234개의 유의어 사전 작성
  - 약어 표현은 풀어서 동일한 의미를 갖게 하였고(ex, AI는 artificial intelligence로 처리), 최대한 띄어쓰기를 고려하였음
- (분석기간) 연도별 분석에서는 전체기간을 '18년 이전, '18년, '19년, '20년, '21년, '22년 상반기 등 6개 기간으로 구분하였음
- (분석도구) 텍스트마이닝을 위해 넷마이너(NetMiner) 4.5 프로그램을 사용
- (분석절차) 연구문제 정의, 기초 데이터 수집, 데이터 전처리, 키워드 분석, 미래 신호 분석, LDA 토픽모델링(토픽도출, 토픽 주제 명명, 토픽별 주요 단어 시각화), 디지털 정책 의제와 디지털 국제규범 비교 등 순차적 연구를 진행

그림 2 텍스트마이닝 연구모형



### 3 디지털 정책 분야 키워드 분석

#### ○ 단어빈도 및 문서빈도 분석

- 단어빈도(TF)와 문서빈도(DF)에서는 data, information, protection, privacy, right 등 데이터 거버넌스와 연관된 단어들에 가장 많이 나타났으며, 다음으로 platform, consumer, content, market, competition 등이 자주 단어로 나타나면서 플랫폼 및 경쟁과 관련된 이슈가 중요하게 거론되는 것으로 유추해 볼 수 있음
  - security, cybersecurity, risk도 많이 등장해 안전·보안 관련 논의도 중심 화두로 대두
- technology, standard 등 디지털 기술 관련 단어와 infrastructure, network 등 디지털 기반 관련 단어들도 높은 순위를 기록하였으며, artificial intelligence는 단어빈도 기준 27위, 문서빈도 기준 29위, 가중치를 고려한 단어빈도 기준 25위를 차지

표 7 상위 30개 기준 단어빈도, 문서빈도, 가중치 반영 단어빈도 분석

순위	TF		DF		TF-IDF	
	키워드	빈도	키워드	빈도	키워드	빈도
1	data	2060	data	664	data	479
2	information	794	information	468	information	241
3	protection	594	protection	394	platform	234
4	platform	552	security	281	consumer	209
5	security	432	platform	272	protection	197
6	consumer	416	consumer	236	cybersecurity	185
7	cybersecurity	335	privacy	219	security	179
8	privacy	332	right	211	privacy	166
9	content	309	technology	194	competition	161
10	right	307	access	189	content	160
11	market	287	market	184	right	154
12	competition	267	content	177	payment	150
13	technology	265	cybersecurity	164	market	147
14	access	246	communication	162	risk	143
15	risk	238	competition	158	google	134
16	payment	230	risk	156	technology	133
17	communication	205	consent	127	access	125
18	processing	180	processing	120	tax	123
19	tax	169	application	115	transfers	111
20	google	165	payment	113	processing	110
21	consent	163	standard	113	advertising	109
22	transfers	157	violation	111	gdpr	108
23	advertising	154	advertising	99	communication	106
24	gdpr	149	transfers	98	product	102
25	application	146	infrastructure	98	artificial intelligence	100
26	product	145	network	96	infrastructure	100
27	artificial intelligence	142	individual	96	consent	98
28	infrastructure	142	purpose	96	social media	95
29	standard	138	artificial intelligence	94	medium	93
30	social media	134	effect	92	revenue	91

\* 주1): TF : Term Frequency, DF : Document Frequency, TF-IDF : Term Frequency-Inverse Document Frequency)  
 \* 주2): TF-IDF 빈도는 TF빈도에 TF-IDF 가중치를 적용한 것이며, TF 순위보다 TF-IDF 순위가 높은 단어는 음영으로 표시

- 연도별로 단어빈도 순위를 살펴보면, '18년 이전에는 internet이 가장 많이 등장하여 망중립성(net neutrality)이 중요하게 거론된 것으로 유추해볼 수 있으며, '18년 이후 분석기간부터는 data가 최빈출 단어로 등장해 디지털 경제 내 데이터 관련 논의가 지속해서 확장되고 있음을 보여줌
- 인공지능은 '18년(6위)과 '19년(10위)에 비교적 높은 순위에 보였는데, 이것은 해당 시기에 인공지능에 대한 사회적 관심이 고조되었고, 한국, 싱가포르 등 주요국 AI 전략이 해당 시기에 나온 것과도 관련이 있음
- 안전·사이버보안과 관련된 키워드인 security, cybersecurity도 30위 안에서 높은 빈도를 나타내고 있으며, platform과 competition의 출현빈도가 꾸준히 상승하여 최근 디지털 경쟁 관련 논의가 강조되고 있음을 반증하고 있음

표 8 상위 30개 기준 연도별 단어빈도 순위 변화

순위	2018년 이전		2018년		2019년		2020년		2021년		2022년 상반기	
	키워드	TF	키워드	TF	키워드	TF	키워드	TF	키워드	TF	키워드	TF
1	internet	27	data	27	data	44	data	123	data	1046	data	800
2	data	20	gdpr	10	transaction	13	information	46	information	431	information	301
3	content	17	cybersecurity	6	cbpr	11	protection	40	protection	346	platform	206
4	net neutrality	10	revenue	6	privacy	11	privacy	33	platform	311	consumer	191
5	right	9	right	6	contract	10	security	32	security	255	protection	188
6	communication	8	artificial intelligence	5	content	9	consumer	28	cybersecurity	204	security	139
7	access	8	content	5	controller	9	content	28	consumer	192	access	125
8	information	8	presence	5	platform	9	platform	24	privacy	173	content	118
9	telecommunication	7	processing	5	protection	9	tax	24	competition	160	market	113
10	network	6	protection	5	artificial intelligence	7	network	18	right	160	privacy	112
11	protection	6	access	4	delivery	7	threshold	18	market	154	right	110
12	application	5	google	4	cybersecurity	6	telecommunication	17	technology	151	cybersecurity	108
13	payment	5	organization	4	ecommerce	6	access	16	payment	142	risk	96
14	basis	4	authorization	3	information	6	right	16	content	132	technology	95
15	technology	4	breach	3	payment	6	communication	15	risk	125	competition	92
16	purpose	4	controller	3	product	6	enforcement	14	communication	114	google	90
17	mobile applications	4	engine	3	right	6	investment	14	tax	103	medium	79
18	tax	4	interest	3	trading	6	market	14	access	91	processing	74
19	discrimination	3	market	3	facebook	5	strategy	14	processing	91	advertising	67
20	development	3	processor	3	organization	5	competition	13	artificial intelligence	88	communication	65
21	privacy	3	safeguard	3	risk	5	consent	13	transfers	87	individual	65
22	request	3	search	3	cookies	4	google	13	gdpr	84	payment	65
23	traffic	3	strategy	3	copyright	4	internet	12	consent	83	consent	60
24	medium	3	subject	3	good	4	payment	12	product	81	social media	59
25	artificial intelligence	3	transfers	3	subject	4	risk	12	infrastructure	80	website	59
26	consent	3	app	2	accountability	3	social media	12	revenue	80	asset	57
27	message	3	device	2	agent	3	application	11	application	79	transfers	56
28	number	3	incident	2	breach	3	assessment	11	standard	78	violation	56
29	establishment	2	individual	2	certification	3	cybersecurity	11	advertising	75	gdpr	54
30	broadband	2	information	2	communication	3	revenue	11	app	70	product	51

\* 주1): Bi-gram 분석을 수행한 키워드는 음영으로 표시하였음

\* 주2): 디지털 정책 문서출현 동향과의 일관성을 유지하기 위해 DPA(2022)에 포함된 자료의 빈도가 낮은 '09년~'17년 기간은 '18년 이전으로 통합하였음



## ◎ Bi-gram 분석

- 단어 및 문서빈도의 분석을 통해 확인된 data, artificial intelligence, security, cybersecurity, platform, competition 등의 중심단어를 대상으로 ‘Bi-gram’ 분석을 시행하여 단어 동시출현 빈도를 측정하였음
- data와 연관 키워드의 동시 출현빈도를 살펴본 결과, data protection(291회), data processing(85회), data security(81회), data transfers(78회) 순이며, 대체로 데이터 보호·관리 관련 키워드의 동시출현이 가장 많이 등장하였음
  - 총괄 영역에서는 data governance, 유형·형태 영역에서는 data subject, 수집·저장 영역에서는 data collection, 가공·저장 영역에서는 data processing, 이전·공유 영역에서는 data transfers, 보호·관리 영역에서는 data protection이 최상위 동시출현 단어로 나타남

표 9 데이터 관련 Bi-gram 분석

순위	총괄(cross-cutting)		유형·형태		수집·저장		가공·처리		이전·공유		보호·관리	
	Bi-gram	N	Bi-gram	N	Bi-gram	N	Bi-gram	N	Bi-gram	N	Bi-gram	N
1	data governance	16	data subject	49	data collection	33	data processing	85	data transfers	78	data protection	291
2	data infrastructure	6	data context	3	data storage	15	data handler	6	data portability	16	data security	81
3	data strategy	3	data information	3	data access	14	data algorithm	4	data sharing	14	data controller	74
4	data platform	2	data russians	3	data cloud	9	data handling	3	data third party	13	data breach	51
5	data artificial intelligence	1	data set	2	data retention	8	data minimization	3	data localization	9	data privacy	46
6	data cooperation	1	data type	2	data individual	5	data analysis	2	data flows	7	data processor	29
7	data economy	1	data activity	1	data consent	4	data communication	2	data consumer	6	data broker	17
8	data partnership	1	data advertising	1	data backup	2	data development	2	data exchange	5	data right	9
9	data stipulates	1	data bulk	1	data generation	2	data activation	1	data exporter	5	data deletion	6
10			data caller	1	data library	2	data aggregation	1	data recipient	5	data organization	5
11			data digitalisation	1	data space	2	data analytic	1	data sale	5	data standard	5
12			data element	1	data warehouse	2	data combination	1	data cross border	3	data child	4
13			data example	1	data addition	1	data developer	1	data market	2	data ethic	3
14			data extent	1	data aggregator	1	data edge	1	data share	2	data opt out	3
15			data face	1	data area	1	data institute	1	data transmission	2	data owner	3
16			data format	1	data correction	1	data intelligence	1	data border	1	data quality	3
17			data geo spatial services	1	data facility	1	data knowledge	1	data contact	1	data tax	3
18			data healthcare	1	data house	1	data machine readable	1	data customer	1	data administrator	2
19			data identification	1	data opt in	1	data mining	1	data delivery	1	data app	2
20			data item	1	data pertaining	1	data minor	1	data export	1	data care	2
21			data loyalty	1	data place	1	data packet	1	data harvesting	1	data content	2
22			data meaning	1	data site	1	data software	1	data link	1	data disclosure	2
23			data message	1					data location	1	data google	2
24			data organ	1					data marketplace	1	data health	2
25			data patient	1					data passing	1	data holder	2
26			data profiling	1					data payment	1	data interface	2
27			data record	1					data publisher	1	data request	2
28			data resource	1					data purchase	1	data residency	2
29			data video	1					data purpose	1	data ability	1
30			data website	1					data retailer	1	data abuse	1

\* 주1): Bi-gram은 Source 키워드(data)와 Target 키워드를 공백으로 연결하여 복합어로 구성한 것임

\* 주2): 음영으로 표시한 부분은 동시출현 빈도가 5회 이상인 것을 의미함

- 인공지능과 동시출현 빈도가 가장 높은 단어는 strategy(7회)로 확인되었으며, 이어서 machine learning(6회), internet of things(4회), data(4회) 순으로 나타나, 국가별 AI 전략 발표와 개발이 확산하는 상황을 대변하고 있음
- security의 동시출현 빈도는 risk(13회), standard(11회), assessment(9회), threat(8회), technology(6회), breach(6회), protection(6회), vulnerability(5회) 순으로 나타나, 국가안보 측면에서 위험, 위협 관련 이슈가 강조되고 있음
- cybersecurity의 동시출현 빈도는 incident(19회), infrastructure(18회), risk(16회), standard(12회), center(10회), audit(5회) 순으로 측정되어 사이버보안 관련 사고, 인프라, 표준, 위험 관련 논의가 중요한 위치를 차지하고 있음
- 최근 3년간 단어빈도 출현이 꾸준히 증가하고 있는 platform의 동시출현 빈도는 youtube(11회), work(9회), accountability(4회), economy(4회), market(4회), worker(4회), news(3회), monopoly(3회) 순으로 나타나, 플랫폼 관련 독점 및 콘텐츠 이슈가 강하게 강조되는 양상임
- competition의 동시출현 빈도는 consumer(21회), market(19회), concern(4회), digital markets(3회), meta(3회), in-app(3회) 순으로 확인되면서 플랫폼 이슈와 결부되어 디지털서비스 시장 내 공정경쟁 이슈가 강조되고 있음

표 10 AI, 보안, 사이버보안, 플랫폼, 경쟁 관련 Bi-gram 분석

순위	Artificial Intelligence			Security			Cybersecurity			Platform			Competition		
	Source	Target	N	Source	Target	N	Source	Target	N	Source	Target	N	Source	Target	N
1	artificial intelligence	strategy	7	security	risk	13	cybersecurity	incident	19	platform	youtube	11	competition	consumer	21
2		machine learning	6		standard	11		infrastructure	18		work	9		market	19
3		internet of things	4		assessment	9		risk	16		accountability	4		concern	4
4		data	4		threat	8		standard	12		economy	4		digital markets	3
5		governance	3		technology	6		center	10		market	4		meta	3
6		product	3		breach	6		threat	8		worker	4		in app	3
7		industry	2		protection	6		audit	5		news	3		internet	3
8		technology	2		vulnerability	5		data	4		monopoly	3		distortion	2
9		cloud	2		standardisation	4		strategy	4		online marketplace	2		network	2
10		algorithm	2		privacy	4		vulnerability	4		website	2		enforcement	2
11		intellectual property	2		resilience	4		program	3		publisher	2		app store	2
12		conference	1		transparency	4		training	3		information	2		innovation	2
13		transaction	1		reliability	4		item	3		liability	2		data	2
14		cross border data transfers	1		data	4		event	3		location	2		watchdog	1
15		risk management framework	1		model	3		artificial intelligence	2		product	2		discrimination	1
16		innovation	1		protocol	3		challenge	2		risk	2		factor	1
17		development	1		exchange	3		profession	2		facebook	2		regard	1
18		risk	1		market	3		certification	2		portal	2		concentration	1
19		center	1		investment	3		expertise	2		bargain	1		high capacity network	1
20		goal	1		telecommunication	3		architecture	2		right	1		barrier	1

\* 주: 동시출현 빈도가 5회 이상인 것은 음영으로 표시하였음

### 3 디지털 정책 분야 미래신호 분석

#### ◎ 분석대상 키워드 선별

- 데이터 전처리 작업을 통해 도출된 2,630개 키워드 모두를 미래신호 분석에 활용할 경우 불필요한 노이즈가 증가할 수 있어 단어빈도 10회 이상, 문서빈도 5회 이상인 545개 키워드를 추출하고, 이 중에서 상위 300개 키워드만 활용하였음
- Yoon(2012) 등 기존 연구를 참고해 최근에 등장 키워드에 더 큰 의미를 부여하기 위한 파라미터인 시간가중치는 0.05로 설정하였음
- KEM는 DoV 기하평균과 평균 단어빈도, KIM은 DoD 기하평균과 평균 문서빈도를 활용해 맵을 구성하였으며, X축과 Y축으로 구성된 4분면에서 신호영역을 구분하는 기준은 기존 연구를 참고해 평균값을 적용하였음<sup>16)</sup>

표 11 평균 빈도 및 평균 증가율(DoV 평균증가율 기준 상위 30개 리스트)

순위	키워드	평균 단어빈도	평균 문서빈도	DoV 평균증가율	DoD 평균증가율
1	meta	23	9	0.925	1.002
2	crypto currency	17	10	0.665	0.496
3	house	33	29	0.653	0.603
4	supervision	12	12	0.653	0.653
5	distribution	13	11	0.616	0.575
6	channel	10	6	0.590	0.418
7	digital services act	8	3	0.561	0.253
8	goal	14	12	0.561	0.514
9	safety	21	15	0.561	0.396
10	asset	34	17	0.530	0.439
11	platform	110	54	0.506	0.503
12	apple	24	9	0.496	0.253
13	power	28	23	0.490	0.411
14	consumer	69	39	0.484	0.309
15	age	10	9	0.478	0.459
16	competition	53	32	0.473	0.329
17	challenge	10	9	0.459	0.418
18	child	30	15	0.459	0.201
19	relation	13	12	0.459	0.459
20	attack	11	8	0.451	0.330
21	minor	21	11	0.449	0.235
22	design	11	10	0.439	0.439
23	party	15	11	0.439	0.372
24	exchange	16	12	0.429	0.496
25	health	26	16	0.429	0.287
26	site	10	6	0.418	0.253
27	web	7	6	0.418	0.346
28	processing	30	20	0.410	0.288
29	address	14	11	0.401	0.382
30	category	12	10	0.396	0.318
평균 (300개 키워드 기준)		18	13	0.071	0.044

16) DoV와 DoD의 기하평균 산출에 대해서는 박찬국 외(2015)를 참고하였음

◎ 미래신호 분석 결과

- KEM에서는 300개 키워드 중 잠재신호가 123개(41%), 약신호가 87개(29%), 강신호가 48개(16%), 강하지는 않지만 잘 알려진 신호가 42개(14%) 순으로 나타났으며, KIM에서도 영역별 키워드 비중이 KEM과 유사하게 나타났음

표 12 미래신호 관련 영역별 키워드 현황(300개 키워드 기준)

구분	KEM		KIM	
	N	비중(%)	N	비중(%)
강신호	48	16	52	17
약신호	87	29	88	29
강하지는 않지만 잘 알려진 신호	42	14	43	14
잠재신호	123	41	117	39
전체	300	100	300	100

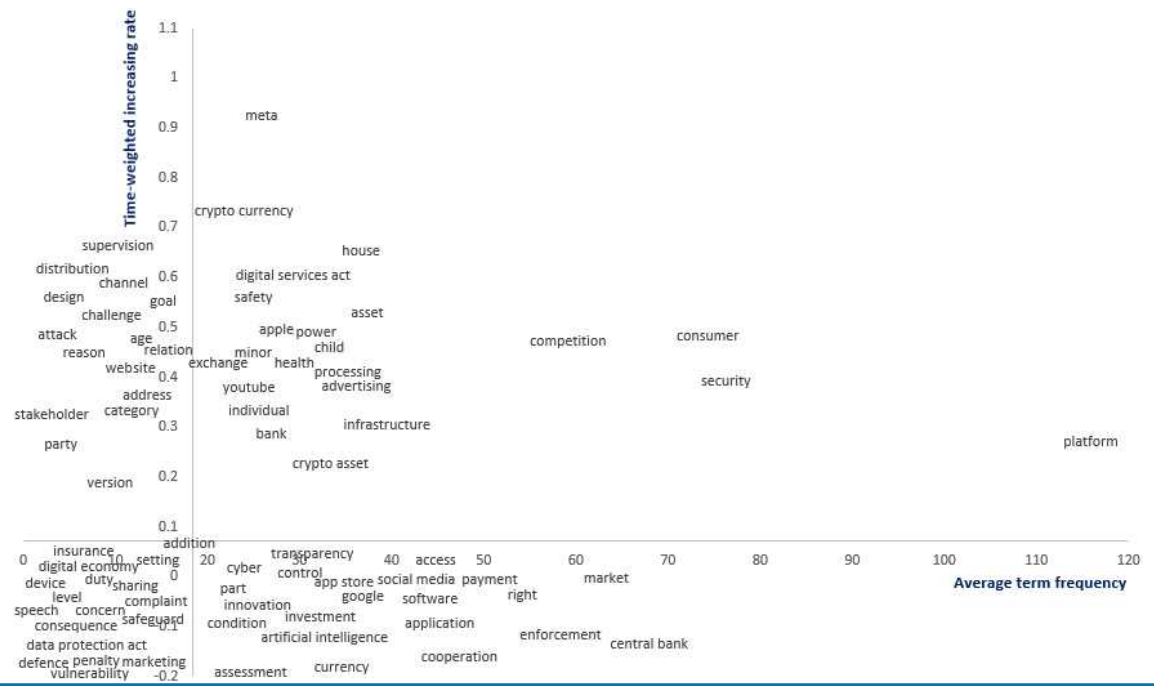
- KEM과 KIM에 공통적으로 잡힌 강하지는 않지만 잘 알려진 신호에는 market, access, software, social media, payment, investment, google, artificial intelligence 등 주로 디지털 규제, 디지털 기반, 디지털 산업 관련 키워드들이 선정
- 강신호에는 safety, platform, power, competition, child, health, security, infrastructure 등 주로 디지털 규제, 디지털 안전 관련 키워드들이 포함
- 약신호에는 cryptocurrency, supervision, distribution, channel, digital services act, design, party, exchange 등 주로 디지털 기반, 디지털 산업 관련 키워드들이 포함

표 13 디지털 정책 분야 미래신호(상위 20개 키워드 기준)

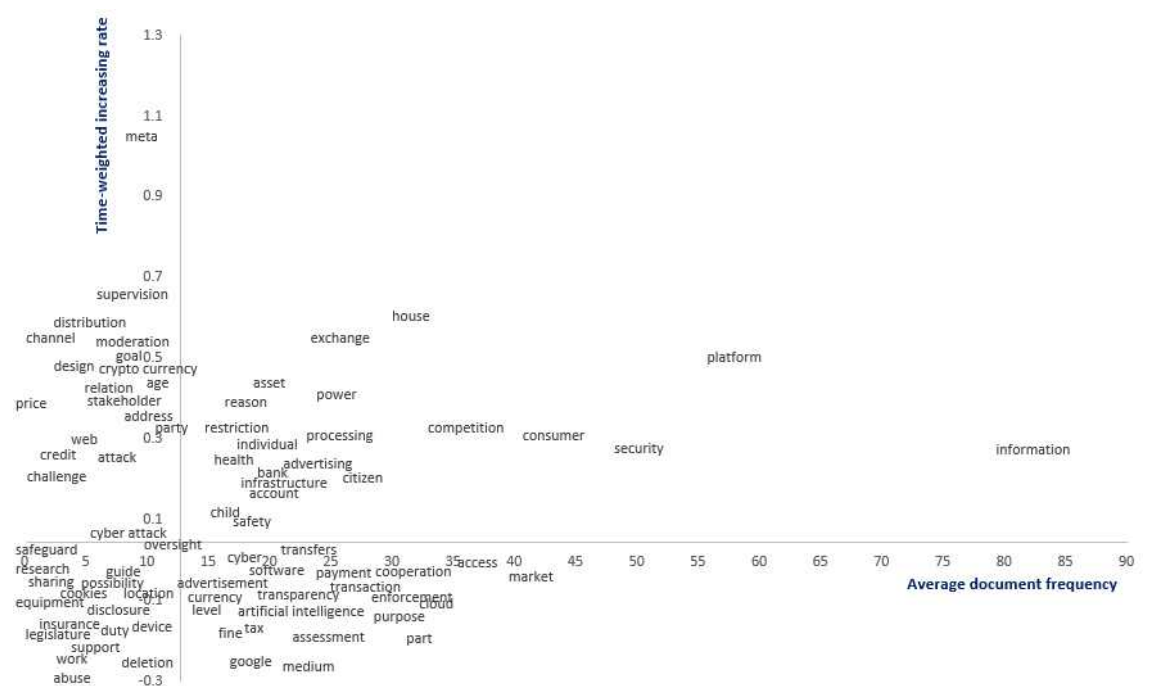
구분	약신호 (weak signal)	강신호 (strong signal)	강하지는 않지만 잘 알려진 신호 (not strong but well known signal)
KEM	cryptocurrency, supervision, distribution, channel, digital services act, goal, age, challenge, relation, attack, design, party, exchange, site, web, address, category	meta, house, safety, asset, platform, apple, power, consumer, competition, child, minor, health, processing, security, advertising, youtube, individual, infrastructure, crypto asset	transparency, market, assessment, currency, innovation, access, software, cyber, social media, control, part, payment, cooperation, app store, right, investment, google, enforcement, artificial intelligence, application
KIM	meta, supervision, distribution goal, crypto currency, exchange, age, relation, moderation, design, channel, challenge, stakeholder, address, party, web, credit, attack, cyber attack, price	house, platform, asset, power, safety, reason, competition, consumer, processing, individual, health, security, restriction, information, advertising, bank, account, citizen, child, infrastructure	fine, market, medium, google, currency, transfers, transaction, assessment, access, transparency, payment, cyber, part, tax, cooperation, enforcement, software, cloud, artificial intelligence, purpose
미래신호	crypto currency, supervision, distribution, channel, digital services act, goal, age, challenge, relation, attack, design, party, exchange, site, web, address, category	house, safety, asset, platform, power, consumer, competition, child, health, processing, security, advertising, individual, infrastructure	transparency, market, assessment, currency, access, software, cyber, social media, part, payment, cooperation, right, investment, google, enforcement, artificial intelligence, application

- \* 주1): KEM : Keyword Emergency Map, KIM : Keyword Issue Map
- \* 주2): TF(단어빈도) 최소 10회 이상, DF(문서빈도) 5회 이상 등장한 키워드만을 고려하였음
- \* 주3): 평균 단어빈도와 가시성(visibility) 정도를 매핑한 KEM과 평균 문서빈도와 확산성(diffusion) 정도를 매핑한 KIM에서 동시에 확인된 약신호, 강신호, 강하지는 않지만 잘 알려진 신호들을 중심으로 미래신호를 도출

**그림 3** Keyword Emergency Map(KEM)



**그림 4** Keyword Issue Map(KIM)



## 4 디지털 정책 분야 토픽모델링

### ◎ 2차 데이터 전처리 및 토픽수 선정

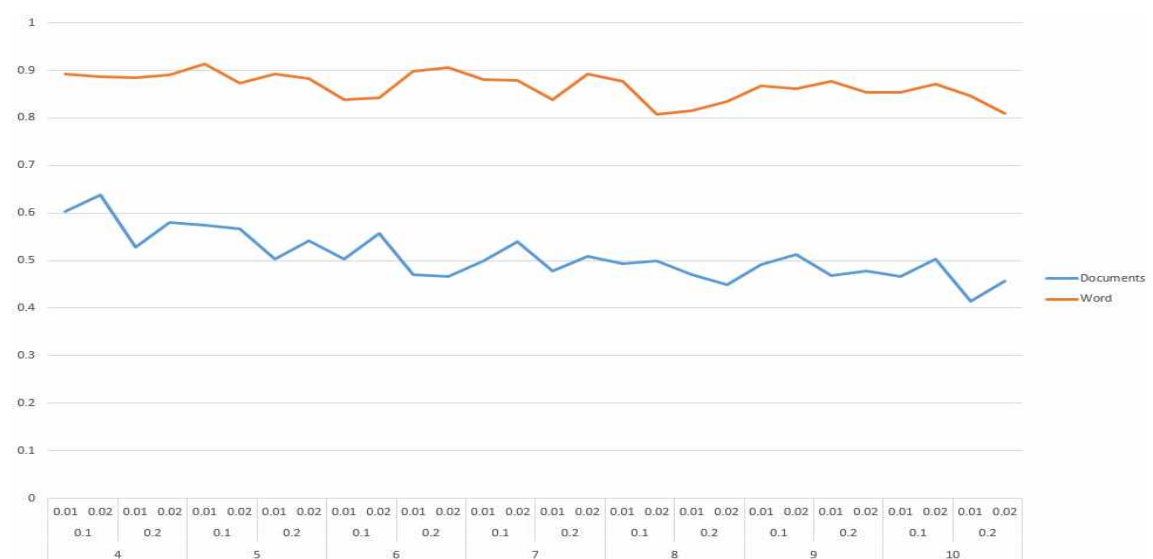
- LDA 토픽분석에 앞서 유의어, 지정어, 제외어 등 사전을 구성하는 작업을 통해 1차 전처리를 마친 단어집합에 대해 TF-IDF 가중치(0.6)를 이용하여 문서마다 빈번하게 등장하는 일반적인 키워드 7개를 제외하는 2차 전처리 작업을 수행
  - 예를 들어, 'data'라는 키워드는 DPA(2022)에 수록된 대다수 문서에 출현하므로 상대적으로 단어빈도가 높은 일반적인 키워드이므로 디지털 정책 분야에서 쟁점화되고 있는 이슈를 파악하기 위한 유의미한 결과도출에는 오히려 지장을 초래할 수 있음

표 14 TF-IDF 가중치 옵션별 제외어 현황

TF-IDF 가중치	제외어(filtered words)
< 0.5	data, information, protection
< 0.6	access, consumer, data, information, privacy, protection, security
< 0.7	access, application, communication, competition, consent, consumer, cybersecurity, data, information, market, platform, privacy, protection, right, risk, security, technology, violation

- LDA 토픽수는 실루엣 계수 평균을 참고해 주제 범주화가 가장 잘 되었다고 판단되는 토픽수를 10개로 설정하였으며, 도출된 토픽의 주제는 주요 단어를 기반으로 명명하였음
  - LDA 분석의 파라미터 입력값은 실루엣 계수 평균의 변화에서 적정 토픽수 10개를 구성하는 파라미터 값( $\alpha = 0.1$ ,  $\beta = 0.01$ , 샘플링 회수 = 1,000회)으로 설정하였음

그림 5 토픽수 증가에 따른 실루엣 계수 평균 변화 추세



\* 주: 세로축은 실루엣 계수 평균, 가로축은 토픽수와 파라미터( $\alpha$ ,  $\beta$ ) 값을 나타냄

## ◎ 토픽모델링 분석결과

- LDA 토픽모델링 분석을 통해 ①플랫폼 중립성, ②인공지능, ③앱마켓 공정성, ④ 알고리즘 중립성, ⑤불법 콘텐츠, ⑥사이버보안, ⑦전자상거래, ⑧개인정보보호 (국경간 정보 이전), ⑨가상자산, ⑩디지털세 등 디지털 정책 분야 의제를 도출
- (토픽1) platform, social media는 모두 플랫폼과 직접 연관된 단어이고, content, communication도 플랫폼을 매개로 구현되며, internet은 플랫폼 관점에서 망중립성을 의미하는 단어인 점을 고려해 토픽1의 주제를 **플랫폼 중립성**으로 명명
  - 인터넷 비즈니스 성장과 함께 다른 활동에 기반 역할을 하는 플랫폼<sup>17)</sup> 유형이 인터넷 접속, 포털, 소셜미디어, 검색엔진, 운영체제, 오픈마켓 등으로 다양화되었고, 플랫폼 사업자의 역할이 중요해지게 되면서 기존 망중립성 의무의 연장선에서 정보차단금지, 비차별, 투명성 등 플랫폼 중립성 의무에 관한 논의가 디지털 분야 규제 이슈로 대두
- (토픽2) artificial intelligence, technology, cloud는 모두 인공지능과 직접 연관되는 단어이고, development, strategy는 인공지능 기술개발을 위한 국가전략을 의미하는 것이므로 토픽2의 주제를 **인공지능**으로 명명
  - 경제 전반의 회복력, 생산성, 성장 및 혁신에 기여한 인공지능 기술의 중요성이 인정되고 있으며, 인력, 데이터, 연산능력, 자금 등 인공지능 경쟁우위 확보를 위한 디지털 산업 정책 및 전략 개발이 글로벌 차원에서 진행되고 있음
- (토픽3) competition, market은 디지털서비스 시장의 경쟁과 맥락이 연결되며, google, app store, app은 구글과 애플의 모바일 앱마켓을 의미하여 토픽3의 주제는 **앱마켓 공정성**으로 명명
  - 구글과 애플의 과점상태에 있는 앱마켓 내 인앱결제 의무화 등 앱마켓 거래의 불공평 행위는 국내·외에서 논의되고 있는 디지털 분야 규제 이슈로서, EU는 앱마켓 거래 관련 공정한 거래를 보장하기 위해 디지털시장법(Digital Market Act)을 제정하였음
- (토픽4) platform, facebook, advertising, advertisement는 플랫폼사업자인 페이스북(현재 meta로 변경)의 수익원인 광고와 연관되며, algorithm은 광고를 구현하는 절차, 방법, 명령어를 의미하므로 토픽4의 주제는 **알고리즘 중립성**으로 명명
  - 플랫폼사업자의 알고리즘을 통해 검색, 구매 등 온라인 활동에서 생성되는 개인의 선택 및 선호에 관한 데이터가 맞춤형 광고 및 과금 등 마케팅 활동에 사용되는 것에 대해 소비자가 알 수 없는 경우 디지털 비즈니스에 대한 소비자 불신 문제가 제기되어 알고리즘 투명성·비차별성 관련 중립성 의무 부과 논의가 디지털 분야 규제 이슈로 논의 중

17) 정찬모(2019), 인터넷 플랫폼 중립성 규제론, 박영사.

- (토픽5) cookies, website는 웹사이트 방문기록을 이용하여 이용자의 브라우징 행동을 불법적으로 추적하는데 활용될 수 있고, child, communication은 쉽게 유해 콘텐츠에 노출될 우려가 있으므로 토픽5의 주제를 **불법 콘텐츠**로 명명
  - 디지털을 활용한 의사소통 및 콘텐츠의 생산·유통·소비 방식의 변화에 대응한 편익은 극대화하면서 가짜뉴스와 유해 콘텐츠에 대한 통제력을 강화하려는 움직임은 디지털 분야에서 논의되고 있는 핵심규제 이슈임
- (토픽6) cybersecurity, risk, incident는 사이버보안과 직접 연관된 단어이며, infrastructure, network는 통신망 인프라의 안전성 측면에서 사이버보안과 연결되므로 토픽6의 주제를 **사이버보안**으로 명명
  - 사이버 공격으로부터 IT 시스템과 중요 콘텐츠를 보호하려는 활동은 지속해서 증가하고 있으며, 러시아-우크라이나 전쟁에서의 사이버전 심화, 데이터센터 화재로 인한 통신망 장애 발생 등으로 디지털 안전 관련 이슈는 최근 더욱 강조되고 있음
- (토픽7) product, ecommerce, application은 음악, 영화 등 디지털 제품(digital products)이 전자상거래 앱을 통해 온라인으로 거래되는 것을 의미하고, number, contract는 전자상거래를 통한 구매계약이 코로나 19로 인해 확대된 것을 유추해 볼 수 있으므로 토픽7의 주제를 **전자상거래**로 명명
  - 디지털화로 인한 제품기획·생산·유통·소비 등 글로벌 가치사슬 변화가 전자상거래 이용률 확대<sup>18)</sup>와 디지털 무역 증가로 이어지면서 국제사회에서 진행되고 있는 전자상거래 원활화를 위한 제도적 기반조성 논의는 디지털 기반 및 디지털 산업 분야의 주요 의제임
- (토픽8) right은 정보주체의 권한을 의미하며, processing, controller는 개인정보의 처리, gdpr은 EU의 일반 개인정보보호법<sup>19)</sup>이므로 이들은 모두 개인정보보호와 연관을 지을 수 있고, transfers는 개인정보의 국외 이전을 의미하는 것으로 판단하여 토픽8의 주제를 **개인정보보호(국경간 정보 이전)**으로 명명
  - 개인정보보호 또는 프라이버시 규제는 데이터의 수집·이용 제한과 관련되어 인공지능 활용력에 영향을 미치게 되는데, 인공지능 산업의 성장과 프라이버시 존중 간 균형을 추구하는 논의는 디지털 기반 관련 최대 이슈이며<sup>19)</sup>, 디지털 국제화 측면에서의 국경간 정보 이전 이슈도 중요한 화두로 이어지고 있음
- (토픽9) payment, transaction은 디지털 거래의 지불수단으로 활용되는 암호화폐와 연관되며, asset, investment, bank는 최근 가상자산이 투자자산으로 거래되는 것을 의미하는 것으로 유추하여 토픽9의 주제를 **가상자산**으로 명명

18) 코로나19가 엔데믹으로 전화되면서 전자상거래 성장세는 둔화되는 추세임. [뱅크오브아메리카](#) 자료를 인용한 Wall Street Journal(2022.11.11.)은 유럽의 의류·신발 매출에서 온라인 비중이 코로나19 팬데믹이 한창이었던 '21년 26%에서 '22년에는 22%로 감소될 전망을 토대로 온라인 구매 증가가 항구적이지 않을 것으로 보고 있음

19) 이종용 외(2019), AI와 국제통상: 디지털 통상규범 현황 및 시사점, ETRI Insight Report, 2019-42, 한국전자통신연구원



- 블록체인·분산원장, 암호화 기술을 활용하는 디지털 금융자산을 총칭하는 가상자산은 사용목적(결제코인, 스테이블 코인 등), 유형(지불형 토큰, 증권형 코인 등) 등에 따라 다양하게 분류되며, 루나-테라 사태, 글로벌 가상자산 거래소 FTX 유동성 사태 등 가상자산 거래시장의 불안정성 문제가 발생하면서 투자자 보호조치의 필요성이 글로벌 차원에서 고조되고 있음<sup>20)</sup>
- (토픽10) tax, revenue, sale, platform은 글로벌 디지털 플랫폼사업자의 매출 또는 수입에 대한 과세이슈와 연관을 지을 수 있으며, threshold는 과세에 적용되는 기준을 의미하므로 토픽10의 주제를 디지털세로 명명
  - 다국적 디지털 플랫폼 기업들의 조세회피를 방지하기 위한 디지털세 논의는 글로벌 이슈로 논의되어 '21년 10월에 글로벌 거대 플랫폼 기업의 이윤에 최소 15% 세금 부과에 대한 국제적인 합의 도출이 이루어졌으며, 세계무역기구(World Trade Organization: WTO) 차원에서는 '22년 6월 제12차 통상장관 회의<sup>21)</sup>에서 차기 회의까지 전자적 전송물(영화, 음악 등 국경간 거래되는 온라인 디지털 콘텐츠)에 대한 관세부과 유예 관행 유지를 결정
- '21년 기준 토픽비중(문서비중)은 가상자산(17.5%), 앱마켓 공정성(12.3%), 플랫폼 중립성(12.0%), 전자상거래(11.5%), 알고리즘 중립성(9.6%) 순임
- '20년과 '21년 사이에 토픽비중이 증가한 주제는 알고리즘 중립성, 불법 콘텐츠, 사이버보안, 전자상거래, 가상자산 등으로 나타났는데, 이것은 해당 분야의 정책 관심도가 타 분야보다 상대적으로 높아진 것을 보여주고 있음

**표 15** 디지털 정책 분야 주요 토픽(의제) 요약

토픽	주제 요약	주요 단어 (상위 5개)	토픽비중 ('21년)	토픽비중 추세 ('20년~'21년)
1	플랫폼 중립성	content, platform, social media, communication, internet	12.0%	감소
2	인공지능	artificial intelligence, technology, development, strategy, cloud	8.1%	감소
3	앱마켓 공정성	competition, market, google, app store, app	12.3%	-
4	알고리즘 중립성	platform, advertising, facebook, advertisement, algorithm	9.6%	증가
5	불법 콘텐츠	cookies, child, website, communication, technology	7.4%	증가
6	사이버보안	cybersecurity, risk, infrastructure, incident, network	7.0%	증가
7	전자상거래	product, ecommerce, application, number, contract	11.5%	증가
8	개인정보보호 (국경간 정보 이전)	right, processing, transfers, gdpr, controller	5.1%	-
9	가상자산	payment, asset, transaction, bank, investment	17.5%	증가
10	디지털세	tax, revenue, sale, platform, threshold	9.5%	-

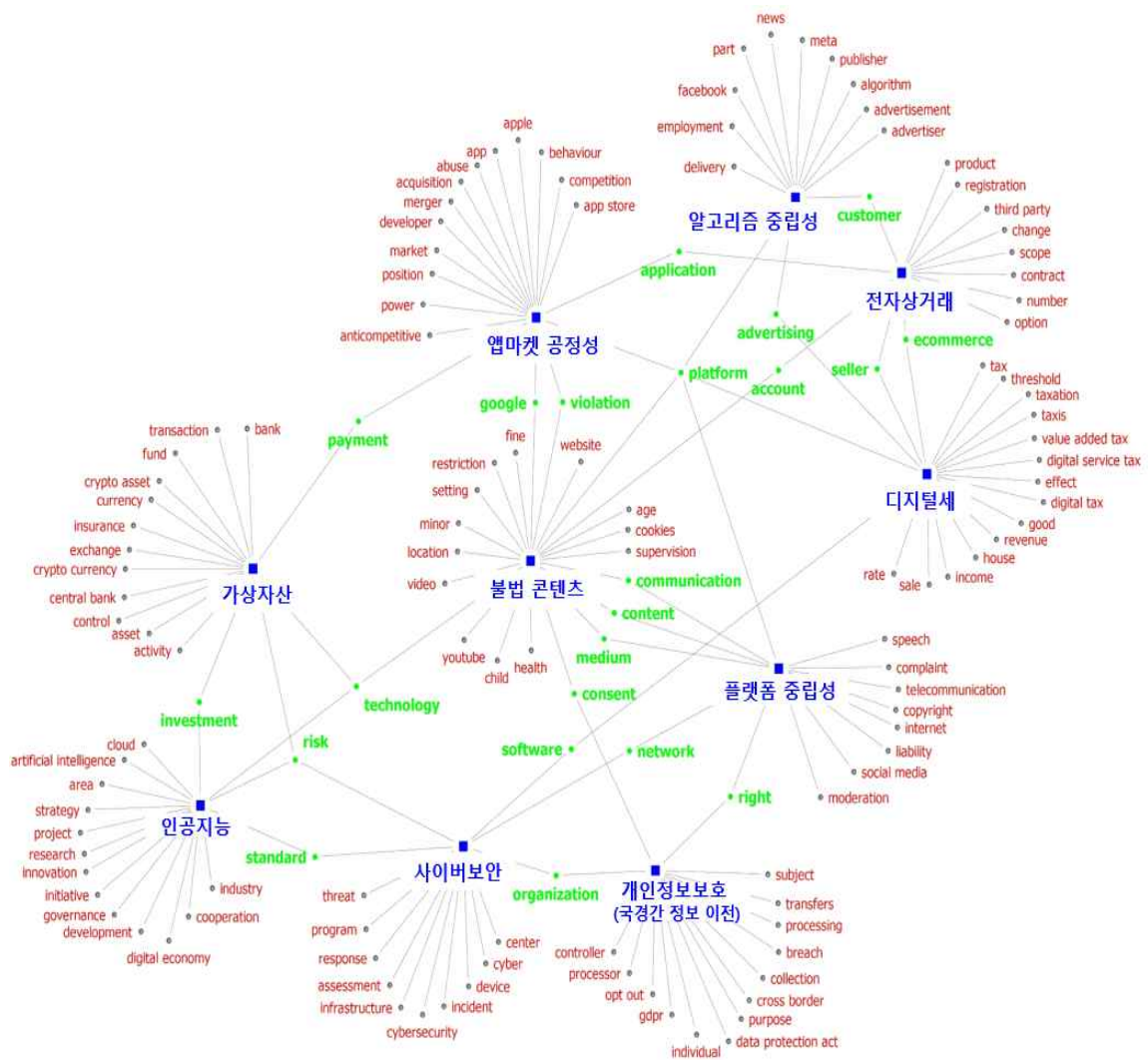
20) 김규림 외(2022), Business Focus 가상자산, 금융 생태계의 새로운 패러다임, 삼성 KPMG 경제연구원.

21) 제12차 WTO 통상장관 회의(Ministerial Conference: MC12)는 당초 '20년 6월 카자흐스탄에서 열릴 예정이었으나, 코로나19 팬데믹으로 연기되어 '22년 6월 스위스 제네바에서 개최되었음

◎ 토픽 상호 간 관계구조

- 토픽별 주요 단어를 시각화한 네트워크를 통해 토픽 간 관계구조를 살펴본 결과, 인공지능 토픽은 가상자산-사이버보안 토픽과는 'risk', 가상자산-불법 콘텐츠 토픽과는 'technology'를 매개로 연결되며, 가상자산 토픽과는 'investment', 사이버보안 토픽과는 'standard'를 매개로 서로 연결되어 있음
- 앱마켓 공정성 토픽은 가상자산 토픽과는 'payment', 불법 콘텐츠 토픽과는 'google', 'violation', 전자상거래 토픽과는 'application'을 매개로 연결
- 'platform'은 앱마켓 공정성 토픽, 알고리즘 중립성 토픽, 디지털세 토픽, 불법 콘텐츠 토픽, 플랫폼 중립성 토픽을 연결하는 핵심 키워드로 역할을 하는 구조임

그림 6 디지털 정책 분야 토픽별 주요 단어 시각화



\* 주: 토픽별 단어 연관 확률이 0.1 이상인 단어를 추출하여 시각화한 것임

- 개인정보보호(국경간 정보 이전) 토픽은 불법 콘텐츠 토픽과는 ‘consent’, 플랫폼 중립성 토픽과는 ‘right’, 사이버보안 토픽과는 ‘organization’과 연결되는 구조임<sup>22)</sup>
- 전자상거래 토픽은 알고리즘 중립성 토픽과는 ‘customer’, 디지털세 토픽과는 ‘seller’, ‘ecommerce’, 불법 콘텐츠 토픽과는 계정을 의미하는 ‘account’를 매개로 연결됨
- 플랫폼 중립성 토픽은 불법 콘텐츠 토픽과 ‘communication’, ‘content’, ‘medium’을 매개로 밀접하게 연결되는 구조를 보여주고 있으며, 사이버보안 토픽과는 ‘network’을 통해 연계되어 있음
- 디지털세 토픽은 온라인을 통해 유통되는 ‘software’를 매개로 사이버보안과 연결되고, 알고리즘 중립성 토픽과는 광고수익 측면에서 ‘advertising’과 연결됨

## 5 디지털 정책 의제와 디지털 국제규범 비교

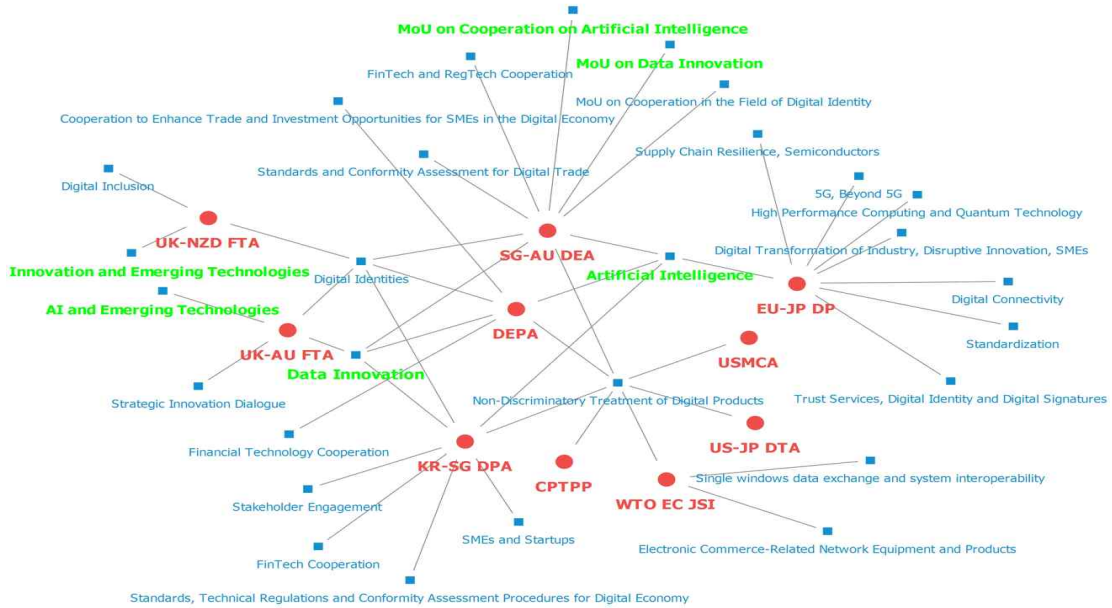
### ◎ 디지털 신기술 관련 국제규범 채택/논의 현황

- (그림 7), (그림 8)과 같이 디지털 신기술 국제규범은 주로 '18년 이후 인공지능, 데이터 혁신, 디지털 신분증, 핀테크, 표준·기술규정·적합성 평가, 중소기업·스타트업, 디지털제품 비차별대우 규범을 중심으로 채택 또는 논의되고 있음<sup>23)</sup>
- 인공지능, 데이터 혁신, 디지털 신원에 관한 국제규범은 UK-NZD FTA, SG-AU DEA, KR-SG DPA, UK-AU FTA, EU-JP DP, DEPA 등 '20년부터 체결된 양자·복수국 디지털 협정에서 협력조항으로 채택
  - 인공지능 규범은 '20년 SG-AU DEA에서 최초로 채택되었으며, 규범내용은 ①AI 기술 채택 및 이용의 중요성 인정, ②AI 기술 사용 관련 윤리적 거버넌스 체계 개발 중요성 인정, ③ AI 거버넌스 체계 개발 및 채택 관련 협력 촉진, ④AI 거버넌스 개발 시 국제적 인정 원칙 또는 지침 고려 등으로 구성
  - 데이터 혁신 규범은 ①국경간 데이터 이동, 데이터 공유 촉진 환경 조성의 필요성 인정, ②규제 샌드박스 활용 데이터 공유 프로젝트 협력, ③데이터 이동성 관련 정책·표준 개발 협력, ④데이터 혁신 관련 연구·산업 사례 공유 등으로 구성
  - 온·오프라인에서 안전하고 편리한 신원증명에 활용되는 디지털 신원 규범은 ①국가 간 협력 중요성 인정, ②공통 표준개발, 법적 효과 인정, 국제적인 프레임워크 개발 지원, 전문지식 교환 등 상호운용성 증진 노력, ③ 당사국의 정책목적 달성을 위한 재량적 조치의 채택 및 유지 허용 등으로 구성

22) 경제개발협력기구(Organisation for Economic Co-operation and Development: OECD)는 ‘온라인 플랫폼’, ‘국경간 정보 이전’, ‘디지털 안전’ 등을 글로벌 디지털 경제를 촉진하는 최우선 정책 의제로 선정하고, 국가 간 분절화(fragmentation)를 방지하기 위한 논의를 추진 중(OECD, 2022b)

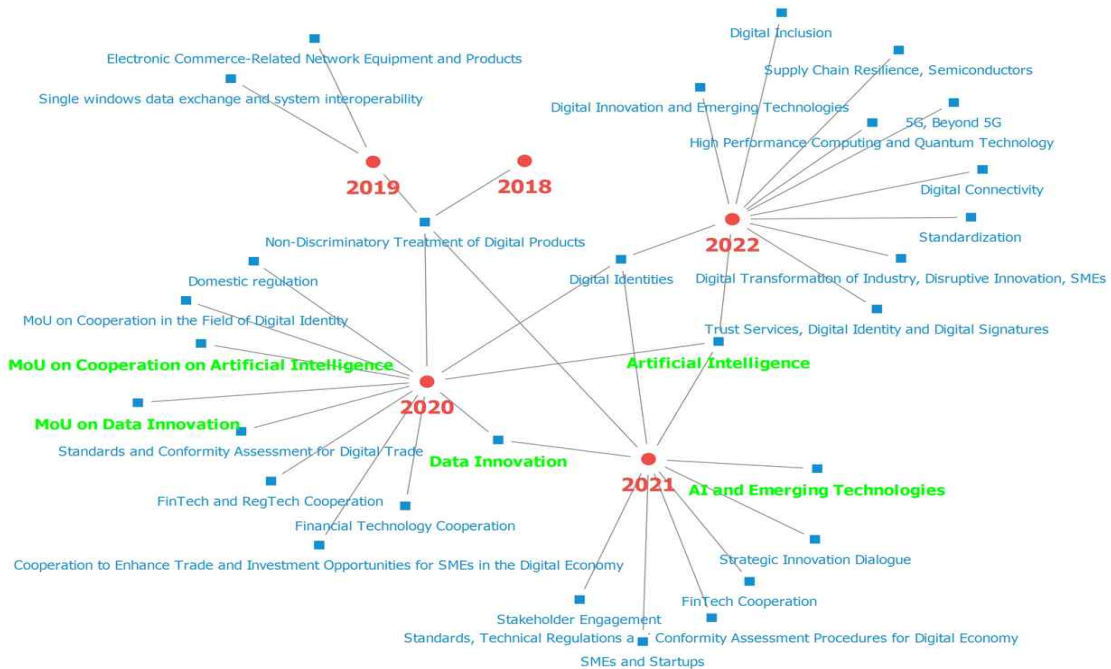
23) 핀테크, 표준·기술규정·적합성 평가 등 기타 디지털 신기술 관련 국제규범에 관한 자세한 내용은 공개된 개별 협정문이나 관계부처 합동(2022)을 참고할 것

그림 7 디지털 신기술 관련 주요 국제규범 현황(협정별)



- \* 주1): ●는 디지털 협정을 의미하며, ■는 채택 또는 논의 중인 디지털 신기술 관련 국제규범을 나타냄
- \* 주2): CPTPP: 포괄적·점진적 환태평양경제동반자협정('18년 발효), SG-AU DEA: 싱가포르-호주 디지털경제협정('20년 발효), USMCA: 미국-멕시코-캐나다 협정('20년 발효), US-JP DTA: 미일 디지털통상협정('20년 발효), DEPA: 디지털경제 동반자 협정('20년 서명), UK-AU FTA: 영국-호주 FTA('21년 서명), KR-SG DPA: 한-싱 디지털 동반자 협정('21년 서명), UK-NZD FTA: 영국-뉴질랜드 FTA('22년 서명), EU-JP DP: EU-일본 디지털 파트너십('22년 서명), WTO-EC JCI: WTO 전자상거래 공동선언문 협상(진행중)

그림 8 디지털 신기술 관련 주요 국제규범 현황(연도별)



## ◎ 디지털 정책 의제와 디지털 국제규범 간 연계성

- 토픽모델링에서 도출된 본 연구의 10대 디지털 정책 의제는 의제별 고유한 특성을 반영해 다음과 같이 분류하였음
  - 플랫폼 중립성, 앱마켓 공정성, 알고리즘 중립성, 불법 콘텐츠 등 4개의 토픽을 디지털 규제 분야의 의제로 분류
  - 사이버보안 토픽은 디지털 안전, 개인정보보호, 전자상거래(전자적 거래 기반조성·신뢰 형성) 토픽은 디지털 기반 분야의 의제로 분류
  - 인공지능, 전자상거래, 가상자산 등 3개의 토픽은 디지털 산업 분야의 의제로 분류
  - 디지털세, 개인정보보호(국경간 정보 이전) 토픽을 디지털 국제화 분야의 의제로 분류
- 한싱 DPA(한국-싱가포르 디지털 동반자 협정, '21년 12월 타결)<sup>24)</sup> 규범을 기준으로 살펴본 결과, 디지털 정책 의제와 디지털 분야 국제규범의 연계성은 존재
  - 디지털 규제 분야 의제는 망중립성 원칙을 반영하고 있는 '전자상거래를 위한 인터넷 접근 이용 원칙' 조항, 경쟁이슈를 다루고 있는 '디지털 경제에서의 경쟁' 조항<sup>25)</sup>, 스팸 메시지를 규제하는 조치를 채택한 '원치 않는 상업용 전자 메시지' 조항과 연계 가능
  - 디지털 안전 분야는 온라인에 대한 신뢰 형성에 기여하는 사이버보안의 중요성과 컴퓨터 보안 사고 대응 역량 구축의 중요성을 인정하는 '사이버보안 협력', '온라인 안전·보안' 조항과 연계
  - 디지털 기반 분야는 '개인정보보호', '데이터 혁신', '정부 데이터 공개' 등 데이터 조항들과 '전자인증·전자서명', '디지털 신원', '디지털 제품 비차별대우', '디지털 경제를 위한 표준·기술규정·적합성 평가절차', '온라인 소비자 보호', '소스코드', '암호기법을 사용한 ICT 제품' 등 전자적 거래의 기반조성 및 소비자·기업 신뢰 형성에 기여하는 조항들과의 직·간접적인 연결고리가 존재
  - 디지털 산업 분야의 인공지능 의제는 '인공지능' 조항, 전자상거래 의제는 '중소기업·스타트업' 조항과 직·간접적 매칭이 가능하며, 가상자산 의제는 '핀테크 협력' 조항과 연계성이 있음
  - 디지털 국제화 분야의 디지털세 의제는 SW 및 기타 디지털 제품 등 무형의 상품에 대한 관세를 부과하지 않는 내용을 담고 있는 '관세(전자적 전송 무관세)' 조항과 연계성이 존재하며, 개인정보보호(국경간 정보 이전) 의제는 민간 영역을 대상으로 국경간 정보 이전을 보장하고, 컴퓨팅 설비의 현지화 요건 부과를 금지하면서 정당한 공공정책 목적 (legitimate public policy objectives)의 달성을 위한 예외조치의 채택·유지를 허용하는 '전자적 수단에 국경간 정보이전' 조항과 '컴퓨팅설비의 위치' 조항과의 연계성이 존재
- 인공지능, 데이터 혁신 등 디지털 정책과 관련 현재의 대다수 국제규범들이 협력·노력조항의 '연성법(soft law)' 단계에 있어 실제법적 의무를 포함하는 수준으로 규정되기까지는 상당한 시간이 필요

24) 한싱 DPA는 암호기법을 사용한 ICT 제품, 소스코드, 디지털 신기술 협력 관련 조항을 포함한 국내 기준 최초 디지털 통상협정이며, SG-AU DEA, DEPA 등 여타 협정과도 비교 가능한 대표적 디지털 국제규범임

25) 현행 자유무역협정에 포함된 경쟁챕터(Competition Chapter)는 투명성, 비차별대우, 절차적 공정성, 공기업에 대한 경쟁법 적용, 경쟁 당국 간 협력 등 경쟁법 집행 관련 원론적 내용을 규정하는 수준이며, 경쟁법 집행상의 결정이 국가마다 다양할 수 있음을 인정하고 있음(정찬모, 2019). 또한, 경쟁정책 이슈는 WTO 차원에서 국제적인 논의가 진행되지 못한 의제에 해당(유지영, 2022)

- 본 연구에서 고려한 한싱 DPA에 포함된 디지털 경제 관련 국제규범 중 당사국 측면에서 구속력이 있는 강행조항은 ‘원치 않는 상업용 전자 메시지’, ‘개인정보보호’, ‘전자인증·전자서명’, ‘디지털제품 비차별 대우’, ‘온라인 소비자 보호’, ‘소스코드’, ‘암호기법을 사용하는 ICT 제품’, ‘관세(전자적 전송 무관세)’, ‘전자적 수단에 의한 국경간 정보 이전’, ‘컴퓨팅 설비의 위치’ 등 10개 조항임

표 16 디지털 정책 의제와 디지털 국제규범(한싱 DPA) 연계

분야	디지털 정책 의제	디지털 경제 국제규범	주요 내용
디지털 규제	플랫폼 중립성, 앱마켓 공정성, 알고리즘 중립성 불법 콘텐츠	전자상거래를 위한 인터넷접근/이용 원칙	• 통상규범 차원에서 망중립성 기본원칙 인정
		디지털 경제에서의 경쟁	• 디지털 시장에서의 경쟁정책 및 법 집행 관련 정보교환 노력
디지털 안전	사이버보안	사이버보안 협력	• 보안사고 대응 역량 구축, 기존 협력 체계 활용, 인력개발 등의 중요성 인정
		온라인 안전·보안	• 다중 이해관계자 접근 중요성 인정 • 글로벌 차원의 이슈 해결을 위한 협력 노력
디지털 기반	개인정보보호	개인정보보호	• 개인정보보호를 위한 국내법 법체계 마련 • 비차별적 개인정보보호 관행 채택 및 정보공표 • 정보교환 노력 및 호환성 촉진 방안 탐색
		전자인증·전자서명	• 전자서명의 법적 효력 인정 • 전자인증 수단 선택권 보장
	전자상거래 (전자적 거래 기반조성·신뢰형성)	디지털 신원	• 디지털 신원 체제의 상호운용성 촉진 및 법적 효과 인정 노력
		디지털제품 비차별대우	• 디지털 제품에 대한 비차별 대우 보장 및 정부 보조금, 방송 관련 적용배제 허용
		디지털 경제를 위한 표준·기술규정·적합성 평가절차	• 디지털 경제 관련 표준 개발 촉진을 위한 국제포럼 참여 및 협력
		온라인 소비자 보호	• 전자상거래 참여 소비자 보호 법제도 채택
		소스코드	• 상용 SW의 소스코드(알고리즘 포함) 이전 및 접근 요구 금지 및 적용배제 허용
암호기법을 사용하는 ICT 제품	• 상용 ICT 제품에 사용된 암호화 기술의 이전 등 부당한 요건 부과 금지 • 사법·행정적 국내법 집행을 관련 적용배제 허용		
디지털 산업	인공지능	인공지능	• 인공지능 거버넌스 틀 개발 관련 협력 촉진 • 인공지능 관련 규제 및 정책 관련 경험 공유
	전자상거래	중소기업·스타트업	• 중소기업·스타트업의 역량 제고를 위한 디지털 기술 활용 정보 및 사례 공유 노력 • 중소기업 스타트업의 온라인 플랫폼 참여 촉진 노력
		데이터 혁신	• 데이터 이동성 정책 및 표준 개발 협력 등 데이터 활용 혁신 지원 노력
		정부 데이터 공개	• 정보 보유 데이터 개방 필요성 인정 • 공공 데이터의 활용 촉진을 위한 협력 노력
가상자산	핀테크 협력	• 당사국 간 핀테크 산업 분야 협력 증진	
디지털 국제화	디지털세	관세	• 전자적 전송(콘텐츠 포함)에 무관세 혜택 보장 • 국내법에 따른 내국세 부과 가능
	개인정보보호 (국경간 정보 이전)	전자적 수단에 의한 국경간 정보 이전	• 사업수행을 위한 국경간 정보 이전 허용 • 정부 정책 권한 허용
컴퓨팅 설비의 위치		컴퓨팅 설비의 위치	• 컴퓨팅 설비의 현지화 요구 금지 • 정부 정책 권한 허용

\* 주: 음영으로 표현된 조항은 구속력이 있는 강행규범

\* 출처: 관계부처 합동(2022) 기반 저자 작성

## IV 맺음말

### 1 결론

- 주요국의 최신 디지털 정책자료를 5개 주제군(디지털 규제, 디지털 안전, 디지털 기반, 디지털 산업, 디지털 국제화)으로 분류하여 디지털 정책문서의 출현 동향과 인공지능 정책 문서출현 동향을 분석하였으며, 텍스트마이닝 기법을 활용해 디지털 정책 분야의 주요 키워드와 미래신호로 잡히는 핵심 키워드를 확인하였음
  - 가장 출현빈도가 많았던 'data' 키워드는 'data protection', 'data security', 'data controller', 'data breach' 등 **보호·관리 관련 키워드**와의 동시 출현이 가장 많이 등장
  - '인공지능'은 사회적인 관심이 고조되었던 '18년과 '19년에 주요 키워드로 등장한 이후 현시점에서는 미래신호 분석에서 '**강하지 않지만 잘 알려진 신호**'로 선정되었음
  - 글로벌 차원에서 국가별 AI 전략의 발표 또는 개발이 확산되고 있는 상황에서 인공지능과 'strategy'의 동시출현이 높게 나타났음
- LDA 토픽분석을 통해 디지털 정책 분야 10대 의제를 실증적으로 도출하였고, 최근 타결된 한싱 DPA의 디지털 경제 국제규범과의 상호 연계성을 확인하였음
  - '인공지능' 토픽은 '21년 기준 8.1%의 비중을 차지하였으며, 최근 '20년~'21년 기준으로는 토픽비중이 감소추세를 보임으로써 인공지능은 이미 보편적인 토픽이 되고 있음을 보여주고 있음
  - 주요 단어 네트워크에서 '인공지능' 토픽은 가상자산과 사이버보안 토픽과는 'risk', 가상자산과 불법 콘텐츠 토픽과는 'technology'를 매개로 긴밀하게 연결된 관계구조를 보였음
  - '인공지능' 토픽은 디지털 경제 국제규범 논의에서 '인공지능'이라는 조항으로 명문화되었으나, 당사국의 노력을 촉구하는 단계에 있음
- 디지털 정책 분야 의제와 관련된 국제협력 및 국제규범 제정 활동은 OECD, WTO, UNCTAD 등 국제기구 차원에서 활발히 전개되고 있으므로 관련 동향을 관찰하고, 다자적 포럼 공간 내 영향력 확대를 지속 추진 필요
  - OECD 디지털 경제정책위원회는 '22년 12월 장관회의에서 온라인 플랫폼, 국경간 정보이전, 디지털 안전, 디지털 격차, 유해 콘텐츠(온라인 허위정보·오정보), 온라인 소비자 보호, 인공지능 및 신흥 기술 관련 신뢰 가능성 증진 등을 세션 주제로 채택하여 디지털 전환의 긍정적 편익과 함께 제기되는 다양한 위험에 대한 국제적 협력 방안을 논의할 예정임<sup>26)</sup>
  - '21년 1월 기준 87개국이 참여 중인 WTO 전자상거래 협상에서는 디지털 환경 하에서의 국제규범 논의가 '18년부터 진행되고 있음<sup>27)</sup>

26) OECD, <https://www.oecd-events.org/digital-ministerial/en/content/documents>

- UNCTAD(United Nations Conference on Trade and Development)는 디지털 데이터의 국경 간 이동에 관한 디지털 경제 리포트('21년), 디지털 통상정책 브리프('22년) 등 자료 발간 및 정기적 회의 개최를 통해 개도국의 디지털 경제 포용성을 촉진하는 활동을 수행 중<sup>27)</sup>
- 디지털 정책 의제와 국제규범 간 긴밀한 연계성은 본 연구의 주목할 만한 포인트이며, 디지털 전략, AI 전략, 디지털 국제규범 등 관련 전략을 유기적으로 연계하는 다양한 접점 발굴은 정책 일관성을 높이고 시너지를 창출함으로써 전략 추진의 전체 효과를 극대화하는 정책과제가 될 수 있음을 제안함
- 디지털 정책 추진의 일관성 측면에서 참고할 사례로서 영국은 AI 분야 강점을 더욱 발전 시키기 위해 정부 내 정책조정을 담당하는 AI 전담 조직을 신설하고 10년 장기 비전에 담긴 국가 AI 전략과 국가 디지털 전략에 국제규범 전략까지 명시적으로 포함하였음<sup>29)</sup>
- OECD는 국가 디지털 전략의 개념적 범위에 포괄적 디지털 전략과 AI, IoT 등 특정 기술에 중점을 둔 조정전략(co-ordinated strategies)까지 포함하여 디지털 전략과 연관된 정책 간 일관성을 강조하고 있으며, 국제규범을 시장 개방성 관련 중요 정책 도메인으로 인식하고 정보이전 및 상호운용성에 대한 고려를 지적<sup>30)</sup>
- '21년 4월 공개된 EU 인공지능법(안)은 가장 높은 수준의 규제가 적용되는 고위험 AI 시스템을 정의하면서 국제규범 관련 전략적 요소를 동시에 고려하였음<sup>31)</sup>
- 디지털 정책의 추진 전략 간 연계 접점으로는 ①국내 데이터 생태계 구축-국외 디지털 인프라 수요 파악, ②중소기업 디지털화-중소기업 해외진출 지원 방안, ③민간기업 AI 개발 지원-디지털 통상협정을 통한 해외 업체와 협력 강화 등이 제안되고 있음<sup>32)</sup>
- 본 연구는 연구 방법론, 연구범위의 확장성, 연구의 시의 적절성 측면에서 다음과 같은 기여점을 제안함
- 첫째 디지털 정책 문서에서 추출한 데이터 기반의 텍스트마이닝이라는 계량적 기법을 활용하여 현재 글로벌 차원에서 논의되고 있는 디지털 정책 의제를 파악하고, 디지털 의제 간 관계구조를 시각화하여 매개 역할하는 핵심 키워드를 실증적으로 확인하였음
- 둘째, 도출된 디지털 정책 의제와 디지털 국제규범과의 연계성을 확인하는 작업까지 연구를 확장하여 디지털 산업 및 규제정책과 디지털 국제규범이 광의적 관점에서 디지털 전환이라는 공통 주제를 통해 상호 간 맥락적으로 연결되는 정책 공간임을 설명하였음

27) [https://www.wto.org/english/tratop\\_e/ecom\\_e/joint\\_statement\\_e.htm](https://www.wto.org/english/tratop_e/ecom_e/joint_statement_e.htm)

28) <https://unctad.org/topic/e-commerce-and-digital-economy>

29) 이종용 (2022), 영국의 디지털 정책: AI와 국제규범 전략을 중심으로, 전자통신동향분석, Vol. 27, No. 6, 한국전자통신연구원.

30) OECD (2020), Going digital integrated policy framework, OECD Digital Economy Papers, No. 292.

31) 유럽위원회는 '21년 4월의 인공지능법(안)에서 공통표준 제정, 적합성 평가 요구 등이 요구되는 고위험 AI 시스템을 EU 시민의 건강, 안전, 기본권에 중대한 피해를 주는 시스템으로 제한하여 국제무역에 대한 잠재적 제약을 최소화하고 있으며, 이보다 더 약한 다른 무역제한 조치가 합리적으로 가용하지 않다는 점에서 부당한 무역 제한을 방지하고 있는 것으로 설명함(EC, 2021)

32) 이규엽 외 (2022), 디지털 전환 시대의 디지털 통상정책 연구, KIEP 정책연구 브리핑, 대외경제정책연구원.



- 셋째, 인공지능, 블록체인 등 디지털 신기술에 대한 국가적 역량이 글로벌 기술패권의 경쟁대상이 된 시점에서 디지털 정책 의제에 관한 실증 연구는 시의 적절성이 있으며, 향후 효과성이 있는 디지털 산업육성 전략뿐만 아니라 인공지능 관련 국제적 논의 참여를 위한 대응전략 마련에 유용한 기초자료로 활용될 수 있음

## 2 향후 연구 방향

- 미래신호 분석의 경우 디지털 정책 분야를 구분하지 않고 분석을 진행하였으나, 후속연구에서 전문가 자문을 통해 키워드를 각 분야별로 살펴보는 미시적 미래신호 분석을 추진하여 시사점을 도출할 필요가 있음
- 기초 데이터로 활용한 DPA(2022)의 고유한 특성으로 인해 미래신호 분석에서 DoV와 DoD 기하평균을 산출하는 과정에 결측치가 존재하는 시점이 제대로 반영되지 못하였다는 점에서 완전한 시계열 자료를 토대로 미래신호를 확인하는 후속 연구를 추진할 필요가 있음
- 토픽모델링 분석에서는 토픽별 키워드에 대한 연구자의 해석을 중심으로 디지털 정책 의제를 도출하였으나, 향후 추진하는 연구에서는 전문가 의견까지 반영하는 과정을 수행하여 토픽 주제를 더욱 객관적으로 명명할 필요가 있음
- DPA(2022)는 지속해서 갱신되는 데이터베이스 자료이므로 토픽모델링 방법론 고도화 작업을 병행적으로 수행하면서 1년~ 3년 주기별 연구를 추진해 동태적 관점에서의 유의미 있는 정책적 시사점을 도출할 필요가 있음

## 참고문헌

### ○ 국내자료

- 관계부처 합동(2022), 한국-싱가포르 디지털동반자협정(DPA) 상세설명자료.
- 김규림 외(2022), Business Focus 가상자산, 금융 생태계의 새로운 패러다임, 삼정 KPMG 경제연구원.
- 김길래 외(2021), Covid-19 대유행 과정에서 나타난 중소기업 이슈 분석에 관한 연구, 중소기업금융연구, Vol. 41, No. 1, pp. 49-77.
- 김성목 외(2020), 텍스트마이닝을 이용한 리빙랩 연구동향 분석, 디지털융복합연구, Vol. 18, No. 8, pp. 37-48.
- 남현동 외(2020), 한국 플랫폼 정부의 방향성 모색 : 공공기관 연구보고서에 대한 토픽 모델링과 네트워크 분석, 디지털융복합연구, Vol. 18, No. 2, pp. 139-149.
- 노설현(2020) 토픽 모델링을 활용한 인공지능 관련 이슈 분석, 디지털융복합연구 Vol. 18, No., pp. 75-87.
- 박대영 외(2021), 토픽모델링 기반의 국내외 공공데이터 연구 동향 비교 분석, 디지털융복합연구, Vol. 19, No. 2, pp. 1-12.
- 박세훈 외(2022), 텍스트마이닝을 통한 디지털 정보격차 현상 연구, 기술혁신학회, Vol. 25, No. 1, pp. 25-42.
- 박찬국 외(2015), 사물인터넷을 통한 에너지산업 발전방향 연구: 텍스트마이닝을 이용한 미래 신호 탐색, 기본 연구보고서, 15-21, 에너지경제연구원.
- 서병조 외(2017), 토픽 모델링을 활용한 한국의 플랫폼정부 연구동향 분석, 정보화정책, Vol. 24, No. 3, pp. 3-26.
- 서호준(2019), 텍스트 네트워크 분석을 활용한 우리나라 과학기술정책 50년의 주요 의제 분석: 과학기술 50년사를 중심으로, 과학기술정책, Vol. 2, No. 2, pp. 171-202.
- 심동녘 외(2020), ICT 정책 지원을 위한 빅데이터 분석 및 예측모형 개발 : 텍스트 자료를 활용한 ICT 이슈 탐지 및 분석 방법론 연구, 경제·인문사회연구회 협동연구총서 20-10-02.
- 우창우 외(2020), LDA 토픽모델링을 통한 ICT 분야 국가연구개발사업의 주요 연구토픽 및 동향 탐색, 한국융합학회논문지, Vol. 11, No. 7, pp. 9-18.
- 유재홍(2022), 디지털 국가의 초석, 디지털 안전, SW 중심사회, 제100호, SPRi.

- 유지영(2022), 글로벌 AI 경쟁과 디지털 통상규범의 의미, KISDI AI Outlook, Vol.8, pp. 1-15, 정보통신정책연구원.
- 이규엽 외(2022), 디지털 전환 시대의 디지털 통상정책 연구, KIEP 정책연구 브리핑, 대외경제 정책연구원.
- 이종용(2022), 영국의 디지털 정책: AI와 국제규범 전략을 중심으로, 전자통신동향분석, Vol. 27, No. 6, 한국전자통신연구원.
- 이종용 외(2019), AI와 국제통상: 디지털 통상규범 현황 및 시사점, ETRI Insight Report, 2019-42, 한국전자통신연구원.
- 정찬모(2019), 인터넷 플랫폼 중립성 규제론, 박영사.
- 조경원 외(2019), 텍스트 마이닝을 이용한 4차 산업 연구 동향 토픽 모델링, 한국정보통신학회 논문지, Vol. 23, No. 7, pp. 764-770.
- 조혜인 외(2019), LDA 토픽 모델링을 이용한 블록체인 학술연구 동향 분석 : 미국, 중국, 한국을 중심으로, 디지털콘텐츠학회 논문지, Vol. 20, No. 7, pp. 1453-1460.
- Cyram(2022), NetMiner 소개와 기초 사용법.

## ○ 국외자료

- Christopher et al.(2008), Introduction to Information Retrieval, Chapter 6, Cambridge University Press.
- DCMS(2022), Digital regulation: Driving growth and unlocking innovation, Policy paper.
- Krigsholm et al.(2019), Applying text mining for identifying future signals of land administration, Land, Vol. 8, No. 181(Nov. 2019), pp. 1-15.
- EC(2021), Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence(Artificial Intellintence Act) and amending certain union legislatives Acts, COM(2021) 2016 final.
- Kubaliya(2016), An introduction to Internet governance, 7<sup>th</sup> edition, DiPLO.
- OECD(2020), Going digital integrated policy framework, OECD Digital Economy Papers, No. 292.
- OECD(2022a), Assessing national digital strategies and their governance, OECD Digital Economy Papers, No. 324(May 2022).

OECD(2022b), Digital enablers of the global economy: Background paper for the CDEP Ministerial meeting, OECD Digital Economy Papers, No. 337(Nov. 2022).

Panichella et al.(2013), How to effectively use topic models for software engineering tasks, ICSE 2013, pp. 522-531.

PwC(2021), Digital opportunities and harms.

Yoon(2012), "Detecting weak signals for long-term business opportunities using text mining of Web news," Expert Systems with Applications, Vol. 39(2012), pp. 12543-12550.

### ○ 웹사이트

Global Trade Alert 홈페이지, <https://www.globaltradealert.org/>

DPA(2022), <https://digitalpolicyalert.org/policy-area>

OECD, <https://www.oecd-events.org/digital-ministerial/en/content/documents>

OECD.AI, <https://oecd.ai/en/dashboards/overview>

Dig.watch, <https://dig.watch/topics>

UNCTAD, <https://unctad.org/topic/ecommerce-and-digital-economy>

WTO E-commerce, [https://www.wto.org/english/tratop\\_e/ecom\\_e/joint\\_statement\\_e.htm](https://www.wto.org/english/tratop_e/ecom_e/joint_statement_e.htm)

### ○ 신문기사

Wall Street Journal(2022.11.11.), End of e-commerce hype train is bad news for planes.

---

## 저자소개

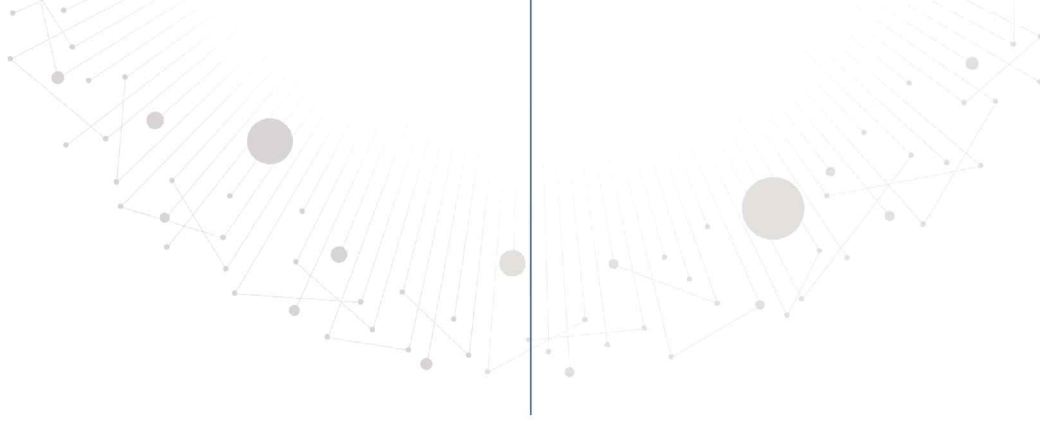
**이종용** ETRI 지능화융합연구소 기술정책연구본부 기술경영연구실 책임연구원  
e-mail: jongyl@etri.re.kr Tel. 042-860-5269

---

## 기술정책연구본부 기술정책 이슈

**발행인** 이 지 형  
**발행처** 한국전자통신연구원 지능화융합연구소 기술정책연구본부  
**발행일** 2022년 12월 31일





[www.etri.re.kr](http://www.etri.re.kr)

본 저작물은 공공누리 제4유형:

출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.



**ETRI** Electronics and Telecommunications  
Research Institute

34129 대전광역시 유성구 가정로 218  
TEL.(042) 860-6114 FAX.(042) 860-6504

