

# 인공지능의 도입원칙



경인교육대학교 사회과교육과  
입법학센터 센터장  
심우민 (법학박사)  
legislation21@ginue.ac.kr  
<http://www.legislation21.org>

# Contents

1. 인공지능의 도입
2. 공공부문 인공지능 도입 유형과 문제점
3. 인공지능 도입원칙
4. 현실적 인공지능 도입원칙(안) 도출
5. 인공지능 도입원칙(안) 설명

# 1. 인공지능의 도입

# 1. 인공지능의 도입(1)

## 인공지능 도입의 영역별 고려(1)

### ▶ 민간영역에서의 인공지능 도입

- ✓ 신기술의 각축장 속에서 민간영역에 도입되는 인공지능의 경우에는 기본적으로 그것의 제공을 통해 수익을 누리는 자, 그리고 그러한 서비스를 이용하는 자들 간의 책임이 가장 중요한 쟁점이라고 할 수 있다.
- ✓ 따라서 공익적 차원에서 민간부분에서 자유롭게 인공지능 기술을 활용하고자 하는 것을 막을 방법도 필요도 없다.
- ✓ 그럼에도 불구하고, 만일 그러한 인공지능의 이용이 공공적, 공익적 차원에서 문제를 발생시킨다거나, 보다 구체적으로는 인권(기본권)에 악영향을 미친다면 이에 대해서는 규범적, 법적 문제가 될 수 있다는 점을 부인하기는 어렵다.

# 1. 인공지능의 도입(2)

## 인공지능 도입의 영역별 고려(2)

### ➤ 공공영역에서의 인공지능 도입

- ✓ 공공영역에서의 인공지능 도입문제는 민간영역에서의 그것과는 달리 신중할 필요가 있는데, 그것의 가장 큰 이유는 공공영역이 가지는 공적 책임과 연관이 되어 있다.
- ✓ 예를 들어, 우리나라의 경우에는 2017년 3월 「지능형 정부 기본계획」을 제시한 바 있는데, 이에 의하면 “지능형 정부” 라는 용어는 지능정보기술을 활용하여 국민 중심으로 정부 서비스를 최적화하고 스스로 일하는 방식을 혁신하며, 국민과 함께 국정운영을 실현함으로써 안전하고 편안한 상생의 사회를 만드는 디지털 新정부를 지향한다는 의미로 사용되고 있다.
- ✓ 이러한 지능형 정부는 사실상 공공영역에서의 인공지능 활용을 보여주는 대표적인 지향점이라고 할 수 있으며, 이러한 기술 및 서비스의 활용은 종래 국가 및 공공기관들이 가지고 있었던 공공성의 수준을 그대로 담지하는 결과에 도달한다.

# 1. 인공지능의 도입 (3)

## 인공지능 도입의 영역별 고려 (3)

### ➤ 민간영역과 공공영역에서의 인공지능 도입 차이

- ✓ 민간영역의 경우에는 사실 기술적 위험이 예견되는 상황 속에서도 그러한 기술 채용을 막을 논리의 구성에 어려움이 있는 것이 사실이지만, 공공영역의 경우에는 전통적인 당해 국가(기관)의 기능과 성격이라는 측면에서 명확한 문제점이 있는 기술의 활용은 다소간의 제재 조치가 불가피할 것으로 보인다.
- ✓ 다만 공공영역의 인공지능 기술 활용도 기본적으로는 관련 기술을 시범적으로 도입해 나갈 필요성은 있는데, 이는 새로운 기술을 활용한 공공성의 창출이 가능한 측면이 분명히 있기 때문이고, 이는 실제로 과거 국가정보화의 맥락도 유사한 견지를 가지는 것이라고 평가할 수 있을 것이다.

# 1. 인공지능의 도입(4)

## 공공영역 도입 원칙에 관한 선제적 고려 고려

- 공공영역에서의 윤리적 관심과 도입원칙 논의
  - ✓ 인공지능 기술 자체는 아직까지 불확정성 (indeterminacy)을 가지는 기술로 평가되고 있다. 이러한 불확정성은 그 기술의 현실화 방향과 내용 자체가 아직까지 명확화 되지 않고 있을 뿐만 아니라, 인공지능의 활용이 일상화·보편화 된 연후에도 실제 그것이 어떠한 영향과 위험을 가지게 될 것인지에 대해서도 정확하게 예측하기 어렵다는 점을 의미한다. 따라서 이러한 기술을 활용하는 데 있어서는 각별한 윤리적 관심이 필요하다.
  - ✓ 특히 공공영역의 경우에는 민간영역 기술 도입의 원칙을 선도 및 유인할 수 있는 성격이 강하기 때문에, 이 영역에 관한 도입 기준 및 원칙에 관한 논의를 선제적으로 고려해 나가야할 필요가 있으며, 민간 영역과는 달리 이러한 도입 원칙에 관한 논의는 과도한 규제 가능성을 그려내는 것도 아니라는 점에서 더욱 의미가 있다.

## 2. 공공부문 인공지능 도입 유형과 문제점



## 2. 공공부문 인공지능 도입 유형과 문제점 (1)

### 공공부문 인공지능 도입 필요성 (Ash Center, 2017) (1)

- 자원의 할당
  - ✓ 행정지원은 과업 완료를 좀 더 빠르게 할 필요가 있음
  - ✓ 문의 응답 시간이 부족한 지원 때문에 길어지고 있음
- 대규모 데이터셋
  - ✓ 데이터셋 규모로 인한 비능률/데이터의 구조화 동향
  - ✓ 내외부 데이터셋 연계를 통한 새로운 가치(인사이트) 창출
- 전문성 부족
  - ✓ 데이터셋 규모로 인한 비능률
  - ✓ 세부 분야는 외부 전문가를 활용

## 2. 공공부문 인공지능 도입 유형과 문제점 (2)

### 공공부문 인공지능 도입 필요성 (Ash Center, 2017) (2)

#### ▶ 예측 가능 시나리오

- ✓ 기존 축적된 데이터를 통해 예측이 가능한 부분이 있음
- ✓ 예측이 필요한 대응에서 신속한 판단의 도움을 받을 수 있음

#### ▶ 절차적 문제해결

- ✓ 공공업무의 성격 중 반복적인 성격을 가지는 것이 다수임
- ✓ 입력 및 출력 자체가 서로 대응하여 대답이 되는 경우가 많음

#### ▶ 데이터의 다양성

- ✓ 데이터는 시각적/공간적이거나 음성/언어 정보를 포함하고 있음
- ✓ 정성적이고 정량적인 데이터를 정기적으로 요약할 필요가 있음

## 2. 공공부문 인공지능 도입 유형과 문제점 (3)

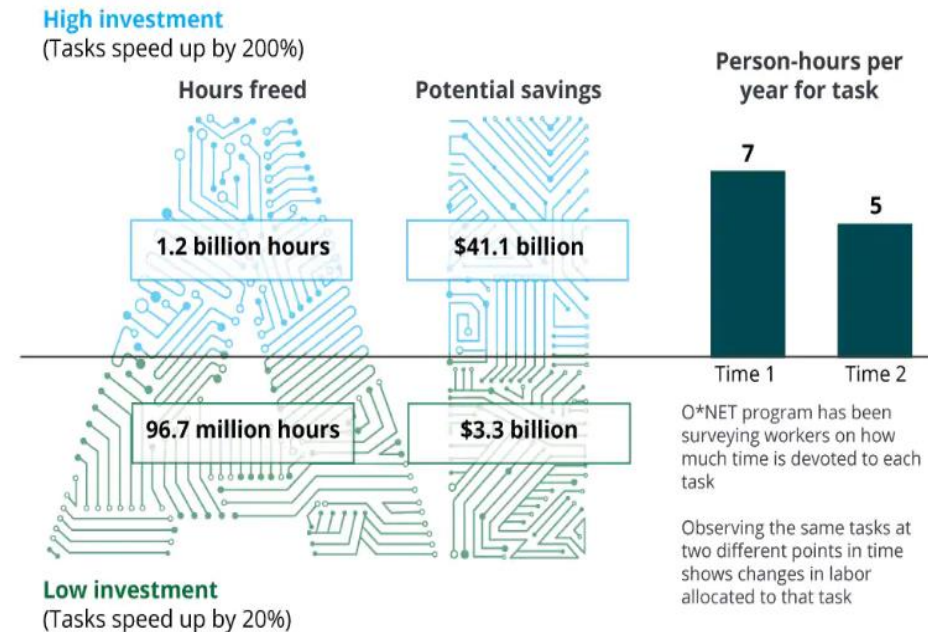
### 공공부문 인공지능 도입 영역 논의(Forbes, 2018)

- 밀려있는 많은 케이스, 요청, 지원 등의 업무에 자동으로 응대하는 ‘봇’ 을 만들어 빠르게 답변하기
- 업무에 적합한 인재 채용
- 자연재해 예측과 구조 노력을 도와주고, 고위험의 업무를 수행하도록 함
- 사이버보안 분야에서 취약성 판단, 위험 예측, 위험 방지와 경감에 활용
- 세무 업무 – 금융 분석, 세금 회피나 이중과세 방지
- 대응 필터링 – 피드백이나 법률 제안 과정에서 방해하는 조작된 반응 같은 노이즈를 필터링해 시민들의 의견이 좀 더 직접 반영되게 할 수 있음
- 돈 세탁, 테러리스트 지원, 사기와 같은 비정규적인 금융 활동 판별
- 인공지능 인력 관리 시스템을 통해 업무 상해 가능성을 사전 예측, 사기성 클레임을 확인,
- 일자리 지원에 대한 분류와 리뷰를 지원할 수 있음
- 인간 행위 흉내 – 고위험 상황에 대한 중요 정보 확인, 화물 배송, 치료 지원 등
- 난해한 업무 – 인공지능 기능이 강화된 드론 등으로 자연재해 지역 조사나 원격의 가스 파이프라인 검사 등을 수행
- 효율 감사 – 정부 기관 내 직원들에 대한 규칙 강화나 도출 행동자 파악 등 자원이 많이 들어가는 영역의 자동화와 빠른 평가

## 2. 공공부문 인공지능 도입 유형과 문제점 (4)

### 공공부문 인공지능 도입 영역 논의 (Deloitte Insights, 2017)

➤ 업무시간 절약을 통한 비용절감



Source: Deloitte analysis.

## 2. 공공부문 인공지능 도입 유형과 문제점 (5)

### 인공지능 도입시 제기가 예상되는 문제점 (1)

- ▶ 신기술 도입시 윤리적 논란
  - ✓ 신기술의 도입은 항상 종래의 규범적 가치기준과 상충하는 부분이 있기 때문에, 윤리적 이슈가 지속적으로 제기될 수밖에 없을 것으로 보인다.
- ▶ 기술 활용 지체와 시민의 기대 수준(cf. 밀레니얼세대의 기계적 대응 선호 가능성)
  - ✓ 신기술을 반드시 정부 서비스에서 활용해야 하는 것은 아니지만, 새로운 신기술 발전에 즉시 대응하지 못하는 경우, 기술적으로 뒤쳐진 서비스 운영이 이루어질 가능성이 높다.
- ▶ 개인 최적화와 프라이버시 문제
  - ✓ 인공지능 서비스는 개인의 취향이나 선호를 더 잘 알수록 개인화된 서비스로 대응 수준이 더 높아질 수 있으나, 이는 개인정보 접근에 대한 이슈를 제기할 수밖에 없다.

## 2. 공공부문 인공지능 도입 유형과 문제점 (6)

### 공공 및 정부영역 인공지능 활용 유형

유형	특징	사례
국민편의 제고	정보제공 및 편의 서비스	지능형 주민센터, 인공지능 행정비서, 지능형 데이터 민관 분석·공유체계 구축, 지능형 객체인식 기반 대형폐기물 처리서비스, 인공지능 및 빅데이터 기반의 지능형 상담시스템, 지능형 문화정보 큐레이팅, 인공지능 기반 개인 수준별 맞춤 교육 서비스, 시민 질의응답 서비스, 시민복지서비스
행정업무 효율화 (과학적 정책결정)	(약) 행정업무 및 판단부담 경감	선제적 행정 프로세스 재설계, 지능형 데이터 민관 분석·공유체계 구축, 상황인지 기반 지능형 안전망, 자가진화형 사이버 안전 기반, 인공지능 및 빅데이터 기반의 지능형 상담시스템, 인공지능 우범화물 검사대상 선별시스템, 지능형 스마트 선별관제 서비스, 로드킬 등 바로신고 서비스 체계, 인공지능을 활용한 과학적 민생치안, 인공지능(딥러닝) 기반의 도로포장파손 실시간 탐지시스템, 음성인식 기반 범정복음 지능형 자동기록 시스템, 인공지능 기반 지중환경오염 예측 및 관리시스템, 지능형 도농복합도시 악취관리시스템, 스마트시티 관리
	(강) 정책결정 직접활용	민의분석 플랫폼, 정책자문관

## 2. 공공부문 인공지능 도입 유형과 문제점 (7)

### 공공 및 정부영역 인공지능 활용 유형별 특성

유형		관련 논제	특성
국민편의 제고		<ul style="list-style-type: none"> <li>• 행정체계 개선(일자리)</li> <li>• 기계적 서비스의 수준</li> </ul>	과거 전자정부 연장 및 변화
행정업무 효율화	(약) 행정업무 및 판단부담 경감	<ul style="list-style-type: none"> <li>• 행정체계 개선(일자리)</li> <li>• 윤리 및 프라이버시</li> </ul>	과거 전자정부 연장 및 변화
	(강) 정책결정 직접활용	<ul style="list-style-type: none"> <li>• 윤리 및 프라이버시</li> <li>• 행정행위의 원인 설명</li> </ul>	새로운 패러다임

## 2. 공공부문 인공지능 도입 유형과 문제점 (8)

### 공공 및 정부영역 인공지능 활용 유형 쟁점사안 가중치

	정보 및 편의 제공	행정업무 단순 경감	정책결정 직접 활용
기술도입 신중성	◎	◎	●
다원적 가치	△	◎	●
신속한 대응	◎	◎	●
책임 행정	●	●	●
인간 개입	○	◎	●
서비스 투명성	◎	◎	●
사생활 보호	◎	●	●
편향성 관리	◎	●	●

\* 범례: ● - 매우 중요, ◎ - 보다 중요, ○ - 중요, △ - 보통



# 3. 공공부문 인공지능 도입 원칙

## 2. 인공지능 도입원칙(1)

### 유럽연합(EU)(1)

➤ EURON 로봇 윤리 13원칙(2007)

- ✓ (1) 인간의 존엄과 인권, (2) 평등, 정의와 형평성, (3) 이익과 손해, (4) 종교적 다양성과 다원성에 대한 존중, (5) 반차별과 반낙인화, (6) 자주성과 개인적 책임, (7) 고지에 입각한 동의, (8) 프라이버시, (9) 신뢰성, (10) 연대와 협력, (11) 사회적 책임감, (12) 이익의 공유, (13) 생물권에 대한 책무 등

➤ EU의회 로봇법 결의안(2017)

- ✓ (1) 인간의 안전, (2) 건강 (3) 보안, (4) 자유, (5) 사생활, (6) 무결성, (7) 존엄성, (8) 자기결정권, (9) 차별금지, (10) 개인 데이터 보호, (11)유익, (12) 무해성, (13) 자율성, (14) 정의, (15) 투명성, (16) 예방, (17) 포괄성, (18) 책무성, (19) 안전, (20) 가역성, (21) 이익 극대화 및 피해 최소화, (22) 프라이버시 기반 설계, (23) 킬스위치(오프아웃 메커니즘), (24) 수정가능성, (25) 추적가능성, (26) 투명성, (27) 예측 가능성, (28) 한계 인식 등

## 2. 인공지능 도입원칙(2)

### 유럽연합(EU)(2)

#### ➤ 신뢰가능한 인공지능 윤리 가이드라인(2019)

- ✓ 인간 중심적 AI 보장 : AI는 윤리적인 목적으로 개발, 구축, 활용되어야 하며 기본권, 사회적 가치, 선행의 원칙, 무해성의 원칙, 인간의 자율성, 정의, 설명 가능성 등을 반영하여야 한다.
- ✓ 기본권, 윤리 원칙 및 가치에 의거하여 인공지능이 인간과 공익을 위해 미칠 수 있는 영향을 사전에 평가하여야 한다. 어린이, 장애인, 소수민족과 같은 취약 집단과 관련된 사안이나 고용주와 종업원, 기업과 소비자와 같이 권력 및 정보의 비대칭성이 있는 상황에는 더욱 특별한 주의가 필요하다.
- ✓ AI는 개인과 사회에 실질적인 혜택과 동시에 부정적인 영향을 가져올 수 있다는 사실을 인지하여야 한다.

## 2. 인공지능 도입원칙(3)

### 영국

➤ 데이터 윤리 프레임워크 6원칙(2018)

- ✓ (1) 이용자 수요(명확한 이용자 수요와 공공의 이익에서 시작하라), (2) 법 및 지침 준수, (3) 최소 데이터 활용, (4) 데이터 한계, (5) 기술 역량, (6) 투명성, (7) 책임 등

### 일본

➤ AI 개발 윤리 가이드라인(2016)

- ✓ (1) 연계의 원칙, (2) 투명성의 원칙, (3) 제어 가능성의 원칙, (4) 안전의 원칙, (5) 보안 원칙, (6) 개인정보보호 원칙, (7) 윤리 원칙, (8) 이용자 지원의 원칙, (9) 책임의 원칙

➤ 인공지능 활용에 관한 7대 윤리지침(2019)

- ✓ (1) AI는 인간의 기본 인권을 침해하지 않는다 (2) 충실한 AI 교육 (3) 신중한 개인정보 관리 (4) AI의 보안조치 확보 (5) 공정한 경쟁 환경 유지 (6) 결정과정에 대한 기업의 설명 책임 (7) 국경을 초월한 데이터 이용환경정비 등

## 2. 인공지능 도입원칙(4)

### 「지능정보사회 윤리 가이드라인」(정보문화포럼, 한국정보화진흥원, 2018)

- 공통원칙 (cf. 세부지침 - 개발자, 공급자, 이용자 별)
  - ✓ ① 공공성 (Publicness)
  - ✓ ② 책무성 (Accountability)
  - ✓ ③ 통제성 (Controllability)
  - ✓ ④ 투명성 (Transparency)

## 2. 인공지능 도입원칙(5)

### 도입 및 활용 원칙 쟁점의 정리 종합

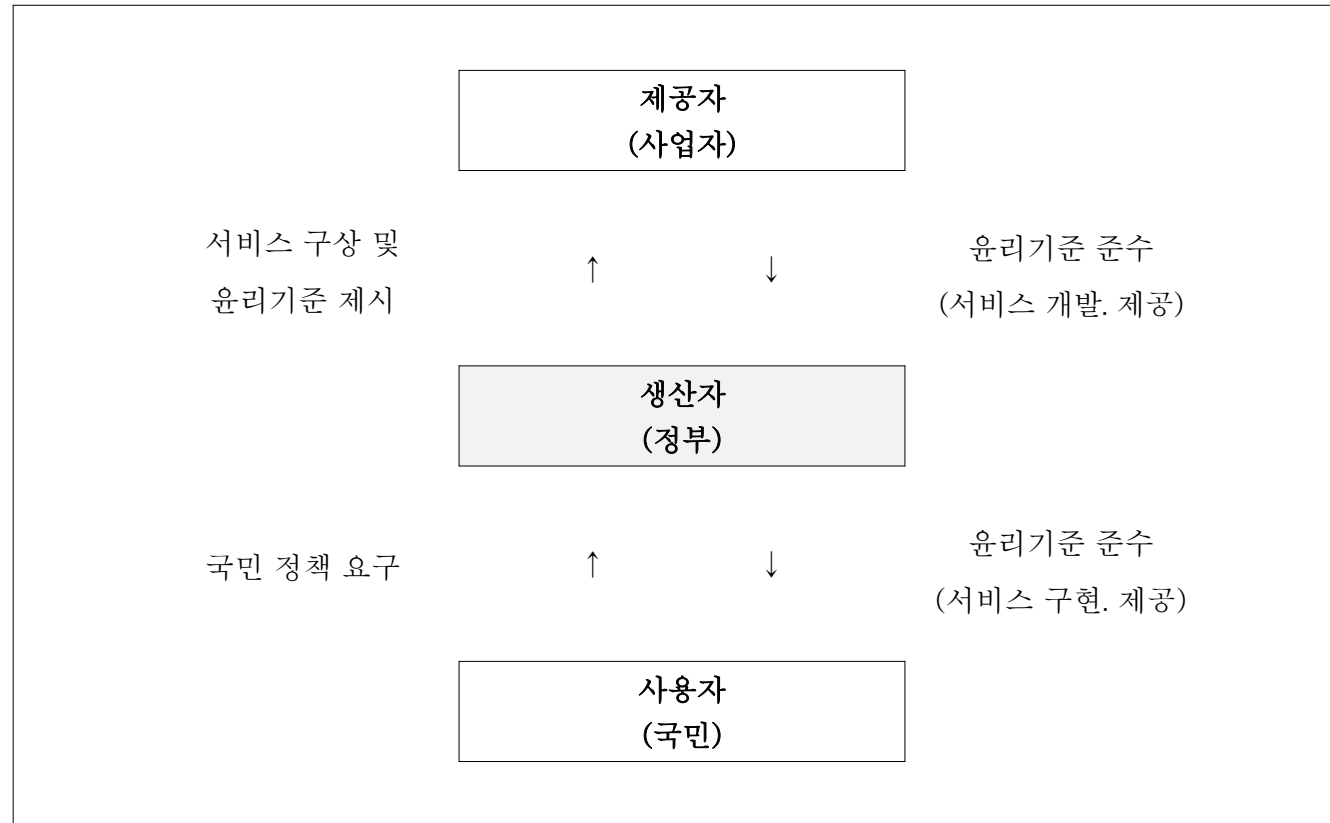
용어 정리		정보 및 편의 제공	행정업무 단순 경감	정책결정 직접 활용	
기술도입 신중성	→	공익성	◎	◎	●
		다원적 가치	△	◎	●
		신속한 대응	◎	◎	●
책임 행정	→	책무성	●	●	●
인간 개입	→	통제성	○	◎	●
서비스 투명성	→	투명성	◎	◎	●
사생활 보호	→	프라이버시	◎	●	●
편향성 관리	→	차별금지	◎	●	●

\* 범례: ● - 매우 중요, ◎ - 보다 중요, ○ - 중요, △ - 보통

# 4. 현실적 인공지능 도입원칙 (안) 도출

## 4. 현실적 인공지능 도입원칙(안) 도출(1)

### 인공지능 도입원칙 전제로서의 행위 규율 관계(1)





## 4. 현실적 인공지능 도입원칙(안) 도출(2)

### 인공지능 도입원칙 전제로서의 행위 규율 관계(2)

#### ➤ 규율관계 분석

- ✓ 실질적으로 공공부문에서의 인공지능 활용 윤리기준을 정립한다고 했을 때, 이것이 직접적으로 적용되는 것은 생산자 영역이다. 즉 관련 서비스를 구상 및 운영하는 주체가 이러한 윤리기준을 실현하는 데 있어 가장 중심적인 지위를 가진다.
- ✓ 제공자의 경우에는 생산자가 발주한 연구개발 요청사항들을 반영하여 수행하고, 윤리적 기준 준수 여부는 종국적으로 생산자들이 판단하게 된다. 물론 제공자들의 경우에는 일반적인 연구개발자 등에게 요구되는 공학 윤리의 준수가 기본적으로 요청되기는 할 것이다.
- ✓ 사용자인 국민에 대해 공공부문의 인공지능을 활용함에 있어 특정 윤리기준 준수를 요청하는 데에는 한계가 있다. 지능형 정부 서비스는 기본적으로 행정적 목적을 위해 활용하는 것으로 국민의 입장에서는 신기술 활용에 대해 특정한 윤리적 의무를 부과하는 것은 타당하지 않은 측면이 있기 때문이다.

## 4. 현실적 인공지능 도입원칙(안) 도출(3)

### 인공지능 도입원칙 및 행위지침(안)(1)

구분	10대 윤리	행위지침
기본 윤리	공익우선	<ul style="list-style-type: none"> <li>공익목적이 명확하지 않을 시 인공지능 활용 자제</li> <li>도입 및 효과에 대한 상시적 법익형량</li> </ul>
	<b>포용성</b>	<ul style="list-style-type: none"> <li>서비스 구상 및 추진상 이해관계자 의견 투영</li> <li>지속적인 이해관계자 서비스 검증 참여</li> </ul>
	<b>응답성</b>	<ul style="list-style-type: none"> <li>국민 입장에서의 기술 혁신성 및 최신성 관리</li> <li>상시적인 위험관리 및 거버넌스 정립</li> </ul>
	책임성	<ul style="list-style-type: none"> <li>법적 책임을 넘어선 국가 책무 고려</li> <li>기속행위와 재량행위 등에 대한 기준 제시(합의)</li> </ul>
	통제성	<ul style="list-style-type: none"> <li>인공지능 및 서비스의 인간적 통제 가능성 확보</li> <li>인간과 인공지능의 상호 주도성 확보</li> </ul>
	투명성	<ul style="list-style-type: none"> <li>시스템 구성·운영 및 알고리즘의 투명성</li> <li>인공지능 활용 여부의 공개</li> </ul>
행정 윤리	프라이버시	<ul style="list-style-type: none"> <li>사생활 관련 정보의 불필요한 수집·추적 제한</li> <li>사생활 관련 정보의 불필요한 집적 제한</li> </ul>
	차별금지	<ul style="list-style-type: none"> <li>데이터셋의 차별 및 편향성 방지 노력</li> <li>인공지능을 통한 차별적 판단결과 공표·활용 제한</li> </ul>
기술 윤리 (R&D 등)	설명가능성	<ul style="list-style-type: none"> <li>인공지능 시스템 설계시 설명가능성 고려</li> <li>시스템에 대한 사후적·기술적 추적 가능성 확보</li> </ul>
	안정성	<ul style="list-style-type: none"> <li>보안 및 프라이버시 관리 가능성 제고</li> <li>오남용 방지 등 서비스의 안정적 제공 기반 마련</li> </ul>

## 4. 현실적 인공지능 도입원칙(안) 도출(4)

### 인공지능 도입원칙 및 행위지침(안)(2)

#### ➤ 기반윤리

- ✓ 첫째, ‘기반 윤리’의 관점에서, 통상적으로 공공부문에서 인공지능 기술을 활용하는 경우에 반드시 선제적으로 고려해야하는 사항들을 제시하였다. 세부적으로는 (1) 공익 우선, (2) 포용성, (3) 응답성, (4) 책무성, (5) 통제성, (6) 투명성을 명시하였다. 이중 포용성, 책무성, 통제성은 우리나라의 정보문화포럼과 한국정보화진흥원에서 제시한 지능정보사회 윤리 가이드라인의 내용과 일치하는 것으로 인공지능 자체의 속성에 기반하여 확보될 필요가 있는 내용들이다. 공공부문에서의 인공지능 활용이라는 점을 감안하여, 공익 우선, 포용성, 응답성을 추가로 제시하였다.

## 4. 현실적 인공지능 도입원칙(안) 도출(5)

### 인공지능 도입원칙 및 행위지침(안)(3)

#### ➤ 행정윤리

- ✓ 둘째, ‘행정 윤리’의 관점에서 고려해야 하는 사항들을 제시하였다. 이는 특히 공공부문(전자정부) 사업 및 관련 정책을 추진하는 담당자들이 견지해야 하는 윤리적 지향점을 제시한 것이다. 여기에는 (1) 프라이버시, (2) 차별금지를 포함하였다. 공공부문에서는 불가피하게 전통적으로 국민의 개인정보 등을 활용할 수밖에 없다는 점에서 프라이버시의 문제가 중시되어 왔으며, 차별 및 편향의 문제는 최근 데이터의 분석적 활용상황에서 강조되고 있는 사안이다.

## 4. 현실적 인공지능 도입원칙(안) 도출(6)

### 인공지능 도입원칙 및 행위지침(안)(3)

#### ➤ 기술윤리

- ✓ 셋째, ‘기술 윤리’ 를 별도로 제시하였다. 이는 기본적으로 공공부문 인공지능 관련 기술을 개발하거나, 그 운영을 지원하는 사업을 수행하는 자들이 견지해야 하는 윤리적 지향점이다. 물론 이러한 기준들은 앞서 기반윤리와 행정윤리의 관점에서 언급한 책무성, 통제성, 투명성 및 프라이버시와 연관성을 가진다. 따라서 이 영역에서 언급하는 윤리 기준은 기술의 개발과 관련하여 강조되어야 하는 시사점을 언급하는 것이다. 이 또한 앞서 언급한 것으로, 여기에는 (1) 설명가능성, (2) 안정성이 있다. 그러나 이 도입기준 자체를 직접적으로 개발자 및 사업자들에게 강제하고자 의도한 것은 아니다. 다소 간접적인 규율로서 공공부문 인공지능 연구개발 및 사업 발주시 관련 공무원(담당자)가 과업지시 및 계약상 이를 포함토록 하는 것이 타당하다.

# 5. 인공지능 도입원칙 (안) 설명

이공지능 도입원칙 (안) 설명

## 5. 인공지능 도입원칙(안) 설명(1)

### 공익우선

- 이 도입원칙(안)에서 언급하고 있는 공익은 단순히 보편적 차원의 공공의 이익만을 의미하는 것이 아니라, 좀 더 구체적으로 인공지능 기술 도입을 위해서는 그것의 공익적 목적이 분명해야 한다는 점을 의미한다.
- 인공지능 기술의 경우에는 다른 기술과는 달리 판단절차 및 결과를 기계적 판단에 의존하는 경향이 있기 때문에, 공익적 목적이 명확하지 않은 가운데 도입될 경우 예측하지 못했던 피해나 의존성이 발생할 가능성이 있다. 따라서 공익적 목적이 명확하지 않은 경우에는 가급적 인공지능의 활용을 자제할 필요가 있다.

## 5. 인공지능 도입원칙(안) 설명(2)

### 포용성

- 알고리즘은 실제 그 설계자의 의도와 학습 데이터의 종류에 따라 가치 편향적 속성을 나타낼 수 있다. 따라서 서비스 구성 및 추진상에 있어 관련 이해당사자들이 심도 있는 참여를 통해, 합의될 수 있는 가치적 기준들을 모색하고 또한 수정해 나가야 할 필요성이 있다.
- 여기서 이해당사자는 기본적으로 이용자인 국민뿐만 아니라, 이들과 이해관계를 가질 수 있는 자들을 의미하며, 더 나아가서는 관련 서비스 정책담당자 및 서비스 구현을 위한 시스템 개발자까지도 포함하는 것이 타당하다.
- 특히 지능형 정부 서비스와 관련성을 가지는 이해관계자들의 참여는 일회성에 그치는 것이 아니라, 당해 서비스가 온전하게 운영되고 있는지에 대해 지속적인 검증절차를 수행하는 과정에서도 이루어져야 한다(이해관계자의 사후적, 지속적 서비스 검증 포함).



## 5. 인공지능 도입원칙(안) 설명(3)

### 응답성

- ▶ 응답성(responsiveness)은 기존에 행정윤리라는 측면에서 지속적으로 제시되어 온 용어이다. 이는 행정체제가 환경의 변화에 얼마나 민감하게 반응하는지 여부를 나타내는 용어로, 상황변화에 따른 정책 수요자들의 요구에 행정의 부합하는지 여부가 핵심 논제라고 할 수 있다.
- ▶ 공익적 목적과 그 도입의 효과가 분명한 경우에 수용 가능한 최선의 수준에서 혁신적이고 최선성을 가지는 기술을 도입할 필요가 있다. 이를 위해서는 현재의 인공지능 기술 활용 가능성에 대한 지속적인 평가가 수반되어야 할 필요가 있다.
- ▶ 응답성의 요청은 공공의 목적을 가지는 전자정부 서비스에 있어 영향평가가 상시화 되어 엄밀한 상황판단에 활용될 필요성을 의미한다. 이는 다수의 인공지능 윤리 가이드라인에서 사회적 영향 평가 등을 강조하는 맥락과 유사한 것이다.

## 5. 인공지능 도입원칙(안) 설명(4)

### 책무성

- 책무성이라는 개념도 최근 각종 인공지능 윤리 가이드라인에서 강조되고 있다. 일반적으로 책무라는 용어는 법적인 책임과는 구별되어지는 관점에서 활용된다. 즉 인공지능으로 인한 역기능과 영향이 아직까지 명확하지 않은 상황에서는 이에 대한 법적 규제와 책임 부과도 명확하지 않기 때문에, 다소 윤리적인 관점에서의 책무를 강조하고 있는 것이다.
- 인공지능을 통한 정책결정으로 인해 발생할 수 있는 공법적 문제점에 대해 고려할 필요가 있다. 인공지능을 통한 정책결정의 대상이 기속행위인 경우에는 판단 알고리즘만 적절히 구성되어 있다면 문제의 소지가 크지 않을 수 있다. 그러나 재량행위에 관해 인공지능 알고리즘을 활용하고자 하는 경우에는 어느 수준까지 인공지능을 활용할 것인지, 그리고 인간의 판단은 어떠한 식으로 개입하는지 여부에 대해 명확한 기준을 제시 및 공표할 필요가 있다.

## 5. 인공지능 도입원칙(안) 설명(5)

### 통제성

- 인공지능이 종국적으로 자율적인 판단을 지향한다는 점에서, 판단 절차와 기준에 관해 인간이 개입해 들어갈 여지가 줄어들 가능성을 예견해 볼 수 있다. 그러나 인간이 개입 없이 전자정부 서비스가 이루어지는 상황에서 문제가 발생하는 경우, 인간이 당해 서비스 및 시스템 운영에 개입할 수 있는 방식을 사전에 마련해 둘 필요가 있다.
- 이러한 통제 가능성을 확보하기 위해서는 당해 서비스 및 시스템 운영에 있어 인간과 인공지능 간의 상호 주도성(mixed initiative)이 확보되어야 한다. 이는 인공지능을 통해 서비스가 이루어지다가 인간이 개입하여 서비스 업무를 수행하더라도, 이 두 서비스 간에는 가급적 질적인 차이가 발생하지 않도록 기술적 조치를 마련해 둘 필요가 있다는 것을 의미한다.

## 5. 인공지능 도입원칙(안) 설명(6)

### 투명성

- 투명성은 관련 시스템과 알고리즘의 구성 과정과 판단 로직, 그리고 그것의 영향 등에 대해 이해 관계자들에게 공개하여, 이를 근거로 이용자들의 판단 근거로 활용될 수 있도록 하는 것이다(이용자 또는 국민의 자기결정권).
- 이와 관련해서는 당해 공공부문 서비스에서 인공지능을 활용하고 있는지 여부에 대해 공개하는 방식을 진지하게 고민해야 할 필요가 있다. 인간과 유사한 행위 및 사고방식을 가지는 기계적 지능을 활용하는 경우, 책임 등의 다양한 문제점들에 대응하는 데 있어 자신이 이용한 서비스에서의 판단이 인간의 것인지 아닌지 여부가 중요해질 수 있다. 따라서 이용자들에게 인공지능 기술의 활용 여부를 공표할 수 있는 방안을 모색해야 한다.

## 5. 인공지능 도입원칙(안) 설명(7)

### 프라이버시

- 빅데이터 분석 기술 및 인공지능 기술의 활용이 증가하면서, 식별 정보와 그렇지 않은 정보를 구분하는 것이 의미가 없어졌다고 볼 수 있다. 따라서 윤리적 측면에서는 관련 정보의 활용이 과연 프라이버시 또는 사생활을 침해하는 것인지 여부에 초점을 맞추어, 지능형 정부의 인공지능 서비스를 구상 및 운영할 필요가 있다.
- 개인 맞춤형 서비스 및 사회적 수요 파악을 위하여 다양한 민간 데이터를 활용하는 과정에서 불필요한 정보까지 수집한다든가, 무단으로 특정인의 정보를 수집하여 그의 행태를 분석 및 추적한다든가 하는 행위는, 명확한 법률적 근거 등이 없는 것이라면 자제되어야 한다. 또한 서비스상의 편의성을 이유로 막대한 양의 사생활 정보를 집적하는 방식의 접근도 제한되어야 할 것이다.

## 5. 인공지능 도입원칙(안) 설명(8)

### 차별금지

- 인공지능 학습 데이터의 편향의 문제는 이제 고전적인 논제가 되었다. 차별이나 편향은 그 자체로 헌법(또는 법률)상 평등권의 침해라고 볼 수도 있고 그렇지 않은 경우도 있을 수 있다. 그러나 그 합법과 불법의 경계는 다분히 맥락적이라고 할 수 있기 때문에, 차별과 편향을 방지하기 위한 노력을 경주해야 할 필요가 있다.
- 데이터셋 자체의 차별성과 편향성을 방지하고자 노력한다고 하더라도, 인공지능이 자체적으로 판단한 결과 자체가 종래 사회적 관념에서 볼 때, 편향성을 가지는 경우가 발생할 수 있다. 이러한 차별성과 편향성의 우려가 있는 경우에는 그러한 판단결과를 활용하거나 공표하는 데에는 주의 기울여야 할 필요가 있을 것이다.

## 5. 인공지능 도입원칙(안) 설명(9)

### 설명가능성

- ▶ 만일 인공지능의 판단결과에 대해 국민에게 합리적으로 설명할 수 없는 것이라면, 그러한 판단결과를 행정 서비스에 활용해서는 안 된다. 이는 결국 인공지능 기술을 활용한 지능형 정부의 인공지능 서비스의 개발을 위한 연구개발 및 위탁 사업을 수행하는 과정에서, 관련 사업의 담당자(공무원)은 설명 가능성의 요구를 명확하게 관철시켜야 한다는 점을 의미한다.
- ▶ 일반적인 인공지능 판단결과에 대한 합리적 설명 가능성의 문제를 넘어서서, 공공부문 인공지능 서비스의 운영 결과가 모종의 손해를 발생시킨 경우 시스템에 대한 사후적·기술적 추적 가능성을 확보할 수 있는 설계를 개발자 및 사업자들에게 요구할 필요가 있을 것이다.

## 5. 인공지능 도입원칙(안) 설명(10)

### 안정성

- 안정성의 요청은 전자정부 서비스와 같은 공공부문 인공지능 서비스가 갖추어야 하는 요소라고 할 수 있다. 특히 정보 및 데이터의 활용은 항상 그 침해에 대비해야 하는 속성을 가진다. 막대한 양적 속성과 민감한 질적 속성을 가지는 데이터의 빈번한 활용이 전제되어 있는 인공지능 활용에 있어서는 과거보다 더욱더 보안과 프라이버시의 쟁점이 부각 될 수밖에 없다.
- 안정적인 지능형 정부의 인공지능 서비스를 제공하기 위해서는, 인위적인 데이터 조작이라든가 오류정보의 제공 등을 방지할 수 있는 기술적 조치들도 갖추도록 할 필요가 있다. 이러한 외부적인 오남용을 방지할 수 있을 때 안정적이고 신뢰 가능한 전자정부 서비스가 이루어질 수 있을 것이기 때문이다.



들어주셔서  
감사합니다.