사람 중심 AI 윤리:
2022년부터 2024년까지의 AI 윤리·신뢰성 포럼

이현경 AI 윤리·신뢰성 포럼 간사, 정보통신정책연구원 부연구위원

1. 머리말

2020년 12월, 대통령 직속 4차산업혁명위원회는 과학기술정보통신부(이하 과기정통부)와 정보통신정책연구원이 마련한 '사람이 중심이 되는 『인공지능(AI) 윤리기준』'을 심의·의결했다. 당시 AI 관련 기반 기술이 급속도로 발전하면서, 각국 및 국제 거버넌스는 앞 다퉈 AI 윤리 원칙 및 가이드라인을 발표했다. 생태주의 철학자 한스 요나스(Hans Jonas)가 개념화한 것처럼 '윤리적 공백(과학기술의 발달과 이를 따라가지 못하는 기존 윤리와의 간극)'을 채우기 위해서다. 이에 우리나라도 '인공지능 국가전략'에 따른 '사람 중심 AI'을 구현하기 위해 글로벌 기준과의 정합성을 갖춘 범국가 AI 윤리 원칙을 마련한 것이다.

『인공지능(AI) 윤리기준』이 마련된 지 약 3년이 지났다. 2024년 5월엔 EU(유럽연합, European Union)가 전 세계 처음으로 포괄적 AI 규제를 위한 'AI 법'을 통과시켰다. 현재는 윤리 원칙·기준들이 점차 AI를 위한 구속력 있는 법적·제도적 프레임워크로 변화되고 있는 중요한 시점이다. 이러한 과도기적 시기에 사회 구성원의 합의를 도출하고, 새롭게 제기되는 AI 윤리 또는 신뢰성 관련 이슈를 논의·발전시킬 수 있는 논의의 장은 너무도 중요하다.

『인공지능(AI) 윤리기준』이 발표된 다음 해, 과기정통부는 '신뢰할 수 있는 인공지능 실현 전략(2021.5)'를 발표했다. 과기정통부는 해당 전략을 통해 AI 윤리기준의 실천 방안을 구체화하고 민간이 자율적으로 AI 신뢰성을 구축할 수 있도록 지원체계를 마련했다. 당시 AI 윤리 실천에 대한 전 세계적 이행 요구가 늘어나며, 정책적으로 대응할 필요성이 점점 높아지고 있었다. 이에 따라 국내 지원체계에 대한 수요 역시 늘어나는 상황을 감안한, 시의적절한 조치로 볼 수 있다.

2021년 당시엔 올해 통과된 EU AI 법 규제안(2021.4)의 초안이 발표됐고, UNESCO(유엔교육과학문화기구, United Nations Educational, Scientific and Cultural Organization) 역시 'AI 윤리 권고(2021.11)'를 발표하면서 각국 정부에 권고 이행을 위한 적극적인 조치를 취할 것을 요청했다.

이렇게 AI 윤리를 실천하는 체계를 마련하면서도, 한편으론 AI에 대한 과도한 규제가 산업 발전을 저해하고 국가 경쟁력을 훼손할 수 있다는 우려 역시 등장했다. 기술개발 목적에 맞는 AI 기술·서비스가 관련 산업 발전을 촉진하면서도, 신기술 부작용을 낮추기 위한 최소한의 안전장치를 확보해야 한다는 것이다. 이에 AI 윤리 체계를 확립하기 위한 정책 거버넌스 구축이 중요해지고 있다. 이번 원고에선 2022년 출범한 AI 윤리·신뢰성 포럼을 소개하면서, AI 윤리정책 거버넌스가 왜 중요한지를 이야기하고자 한다.

2. AI 윤리·신뢰성 포럼 소개

2.1 포럼의 역할

2022년 2월, AI 윤리·신뢰성 포럼이 AI 신뢰성 확보를 위한 거버넌스 구축의 일환으로 마련됐다. AI 윤리·신뢰성 포럼의 목표는 AI 발전과 사회적 확산에 따라 새롭게 등장하는 윤리적 이슈를 논의하며, '신뢰할 수 있는 사람 중심 AI 개발·사용을 위한' 사회적 합의를 구축하는 것이다. 포럼엔 학계, 산업계, 교육계, 법조계, 시민사회 등 분야별 AI 기술·윤리 전문가가 참여했다. 이를 통해 포럼은 AI 윤리 확보를 위한 주요 정책과제에 대해, 각계 전문가와 이해관계자가 의견을 제시하는 창구가 됐다.

포럼의 중요한 기능 중 하나는 정책 결정의 투명성을 달성하는 것이다. 정책 결정자들은 이러한 자리에서 나온 각계각층의 의견을 종합적으로 수렴하고, 미래 정책 방향에 대한 제언을 공식적으로 구하면서 정책 추진과정의 투명성을 확보할 수 있다. 이해관계자들의 폭넓은 의견을 수렴함으로써 정부의 주요 정책 이니셔티브가 다양한 관점을 반영하도록 보장하고, 한편으로 산업계에선 사회적으로 책임 있는 정책의 개발·구현에 도움을 주고자 하는 것이다.

2.1.1 AI 관련 제도적 거버넌스 지원

AI 윤리·신뢰성 포럼을 진행하면서 많이 받는 오해가 있다. AI 관련 법이나 규제를 수립하기 위한 자리로써 포럼이 기능한다는 것이다. 이러한 오해는 AI윤리·신뢰성을 위한 실천 수단이 향후 고착화돼 법적 강제·규제로 작용할 것을 우려하는 것이다.

그러나 AI 윤리·신뢰성 포럼은 제·개정이 경직된 법규범이 가진 한계를 뛰어넘기 위한 것이다. 급격하게 변화하는 기술의 속도에 맞춰 법이나 제도를 계속 변경할 수 없으므로, 각 사회 구성원이 유연하게 대처할 수 있는 환경을 조성하고자 '윤리'의 역할이 필요한 것이다.

최근 기술·서비스의 일상화·보편화로 인해 여러 사회적 이슈가 발생하고 있다. 이에 따라 사회적 우려와 해결 요구가 증대되고 있지만, 앞서 말한 한계로 인해 이를 법·규제로만 해결할 수 없는 상황이다. 때문에 포럼은 이러한 한계를 인정하고, 기업이나 사회 구성원이 자율적으로 AI의 윤리적 활용을 도울 수 있도록 자율적인 실천 방안을 제공하는 데 도움을 주고자 한다.

또 다른 오해 중 하나는 본 포럼에서 'AI의 윤리성'을 이야기할 것이라는 견해다. 강조하고 싶은 것은, 'AI 자체를 자율성을 가진 주체로 인식해 논의하는것'이 아니라는 점이다. 포럼에서 이야기하고 있는 것은, AI를 객체로 이해해 AI을 활용하는 사람들, 또는 AI 기술과 시스템을 개발하고 이용하는 주체들이 가져야 할 윤리 규범이다.

일종의 안전장치로서 윤리 체계를 마련하기 위해선 많은 지원이 필요하다. 여기엔 AI 윤리 확보를 위한 기업의 자율적 활동 지원, AI 윤리에 관한 국내외동향 제공, AI 윤리교육 방향 등 주요 AI 윤리 이슈를 논의하는 것이 포함된다. 이러한 작업들은 법이나 규제 같은 경성 규범은 아니지만, 연성 규범으로서 제도적 거버넌스를 지원해 AI의 신뢰성을 확보한다는 데 그 의미가 있다.

2.1.2 민, 관, 학 논의의 장

학계, 산업계, 시민사회가 함께 참여해 논의하는 공론의 장을 제공하는 것은 포럼의 중요한 기능 중 하나다. 특히 AI 개발과 활용에서 나타나는 여러 이슈는 국제적으로도 국가별·기업별 다양한

이해관계가 첨예하게 얽혀있는 경우가 많다. 이러한 다양한 이해관계를 가진 행위자들이 한자리에 모여, AI 윤리정책에 관한 시각을 공유하는 것 자체만으로도 의미를 가진다.
물론 모든 사안에 있어 참여한 이해관계자들이 공통된 시각을 보이는 것은 아니다. 하지만 입장 차이를 확인하는 것은 다양성을 바탕으로 한 정책 논의에 도움이 된다. 또한 서로 다른 입장에도 불구하고, 거시적 관점에서 필요한 국가 정책의 방향에 대해 고민해 보는 기회가 될 수 있다는 점에서 중요하다.

2.1.3 대국민 정책 의사소통

2022년부터 연중 진행된 포럼의 말미엔 'AI 윤리공개 정책세미나'가 개최됐다. 이는 포럼 운영의 성과와 함께 관련 주요 정책 추진의 결과를 국민과 투명하게 공유하는 자리다. 목표는 주요 이슈에 대한 공론의 장을 제공하고, 윤리적 AI의 개발·활용 필요성에 대한 사회적 공감대를 형성하는 것이다.
실제 공개 정책세미나에선 2022년~2023년의 주요정책 성과물인 AI 자율점검표, AI 윤리영향평가(안), AI 신뢰성 검·인증 방안, 초·중·고·일반인 대상 AI 윤리교육 콘텐츠 등이 소개됐다. 일반 청중을 포함한 참가자들은 이를 바탕으로 한 해의 주요 AI 윤리·신뢰성 정책과제에 대한 결과물을 발표하고 토론을 진행했다.

[그림 1] AI 윤리 확산을 위한 공개 정책세미나

2.2 포럼의 구성

AI 윤리·신뢰성 포럼은 정책 수요를 잘 반영할 수 있도록 매년 그 구성과 성격을 달리해 왔다. 거버넌스 구축에 가장 중요하고도 도전이 되는 과제는 인적 네트워크 형성이다. 어떠한 인적 네트워크를 형성하고 발전시키느냐에 따라, 해당 담론의 내용과 성격이 달라질 수 있다. 제1기부터 제2기까지는 총 3개의 분과를 두었다. 윤리·기술·교육 각 분과에서 적합한 주제를 골라 논의하고, 전체 포럼에서 관련 논의를 공유하는 형태로 포럼이 진행됐다.
올해 제3기에선 세 분과의 구분을 없애고, 하나의 포럼으로 구성됐다. 포럼 구성이나 참여 인원의 변화는 그해의 정책 수요를 반영하는 차원에서 이뤄진다. 포럼의 역할은 사회 전반에 건전한 AI 의식을 확산하고 정책 거버넌스를 구축한다는 큰 목표에서 벗어나지 않는다.

2.2.1 제1기, 제2기 AI 윤리정책 포럼(2021~2023년)

2022년 제1기 AI 윤리정책 포럼에는 총 30명의 위원이 참여했다. 각 위원들은 전문성을 고려해

윤리·기술·교육 분과에 투입됐다. 제1기 위원장으론 고학수 서울대 법학 전문 대학원 교수가 활동했으며, 2022년 10월 개인정보보호위원장으로 취임한 이후엔 공석으로 진행됐다.

1년 동안 활동을 돌아보면, 전체 포럼(공개 정책 세미나)은 4회, 분과회의(윤리 분과회의 3회, 기술 분과회의 4회, 교육 분과회의 3회)는 10회 개최했다. 윤리분과에선 'AI 윤리 체계의 확산' 이슈를, 기술 분과에선 'AI 신뢰성 확보 기술 기반 마련' 이슈를, 교육 분과에선 'AI 리터러시 및 윤리교육 강화' 이슈를 다뤘다.

제1기 포럼엔 네이버, 카카오, LG AI 연구원, 한국마이크로소프트(Microsoft)를 비롯한 국내외 대기업은 물론, 스캐터랩, 셀렉트스타와 같은 스타트업들도 참여했다. 더불어 한국교육개발원, 한국과학창의 재단과 같은 공공기관, 소비자시민모임, 한국소비자연맹 등 시민단체도 참여해 다양성을 확보했다.

2022년 말부터는 오픈 A I(OpenAI)가 개발한 ChatGPT가 대중적으로 이용되면서 초거대·생성형 AI의 사회경제적 영향이 중요하게 조명됐다. 이에 2023년 출범한 제2기 AI 윤리·신뢰성 포럼은 정책 협력 네트워크 구축과 운영에 좀 더 중점을 두고 진행됐다. 전체적으로는 제1기 구성 방식과 유사하게 세개 분과로 운영됐다. 위원장은 김명주 서울여대 정보보호학과 교수가 맡았다.

제2기 포럼에서 참가자들은 AI 윤리에 관한 글로벌 표준화 논의에서 국내 정책을 소개하는 등 적극적인 참여가 중요하다는 것을 재확인했다. 또한 기업 자체의 책임성을 바탕으로 자율규제와 정부 정책 간 조화를 통한 기술 발전이 필요함을 확인했다.

이 외에도 포럼에선 2023년 정부에서 추진한 새로운 디지털 질서 정립과 디지털 권리장전 수립을 위한 주요 논의사항을 다뤘다. 포럼 위원들은 AI 윤리·신뢰성 확보를 위한 디지털 권리장전의 방향에 대해 의견을 제시하는 시간을 마련했다.

한편 포럼과는 별도로 간담회(2023.5)가 개최되기도 했다. 간담회 내용은 기업들의 AI 윤리·신뢰성 정책·사업 추진 현황과 함께, 기업의 자율적인 노력을 공유하는 것이었다. 포럼에는 학계, 교육계, 산업계, 시민단체, 법조계, 공공 등 다양한 분야 인원이 참여하는 만큼, 기업의 활동에만 초점을 맞추기 어려운 점을 고려한 것이다.

2.2.2 제3기 AI 윤리·신뢰성 포럼(2024년)

2024년 4월 출범한 제3기 AI 윤리·신뢰성 포럼은 제1기, 제2기 포럼과는 달리 하위 분과를 나누지 않고 하나의 포럼으로 출발했다. 이는 지난 2년 동안의 포럼 활동에 대한 피드백을 바탕으로 구조적 변화를 꾀한 것이다. 위원장은 이상욱 한양대학교 철학과 교수가 맡았다.

한 포럼 내에서 하위 분과를 두어 운영했을 때의 장점은, 많은 수의 포럼 위원들이 적은 수로 나뉘어 자신의 전문성에 부합하는 분과에서 논의한다는 것이다. 이를 통해 더 심도 깊은 논의를 진행할 수 있다. 반면 전체 포럼 위원들이 별도의 분과 없이 하나의 조직으로 운영되면, 포럼 위원들이 특정 분야에 나눠지지 않고 종합적인 논의와 토론을 할 수 있다. 학계·산업계·법조계·공공·시민사회·국제기구 등 부문별 전문가 총 20명으로 구성된 포럼 위원들은 AI 윤리정책 방향에 대해 논의하고, 국제사회 동향에 따른 우리 정부의 대응, 윤리 정책 실행 수단(AI 윤리 가이드라인, 윤리영향평가, 교육콘텐츠 등) 개선 및 활용도 제고 등을 논의할 예정이다.

제3기 포럼은 앞선 제1기, 제2기 포럼과 비교해 보았을 때, 그 구성과 위상이 달라지고 있다. 지

난 4월 18일 출범한 제3기 AI·신뢰성 포럼은 과기정통부의 민·관 AI 최고위 거버넌스인 'AI 전략최고위협의회'의 윤리·안전 분과로서 기능하게 됐다. 'AI 전략최고위협의회'는 그보다 앞선 4월 4일 과기정통부 장관과 민간을 공동 위원장으로 하여 총 32인으로 출범했다.

여기엔 정책 일반, AI 반도체, 연구·개발, 법·제도, 윤리·안전, 인재 등 AI 분야를 대표하는 민간 전문가 23인과 함께 주요 정부 관계부처 실장급 공무원이 포함돼 있다[1].

AI 거버넌스와 관련, 현재 국제적으로도 여러 이니셔티브가 공존하고 있어 상호운용성 및 정합성 논의가 커지고 있다. 이런 상황에서 AI 관련 국가 정책은 통합적으로 운용될 필요가 있다. 이에 우리 정부도 여러 부처 산하에 있는 AI 관련 협의 거버넌스를 일관성 있게 운영하기 위해 상위 거버넌스를 출범시킨 것이다[2].

이상욱	한양대학교 철학과 교수 **(위원장)**	박성필	한국과학기술원 문술미래전략대학원장
김명주	서울여자대학교 정보보호학부 교수	박찬준	업스테이지 수석 연구원
김경훈	카카오 이사	변순용	서울교육대학교 윤리교육과 교수
김도엽	김·장 법률사무소 변호사	손지원	오픈넷 변호사
김동환	포티투마루 대표	송대섭	네이버 이사
김유철	LG AI연구원 전략부문 부문장	신준호	한국정보통신기술협회 AI신뢰성센터장
김은영	유네스코한국위원회 유네스코의제정책센터장	안성원	소프트웨어정책연구소 AI정책연구실장
노태영	김·장 법률사무소 변호사	이영탁	SK텔레콤 성장지원실장
문명재	연세대학교 행정학과 교수	이화란	네이버 퓨처 AI 센터 리더
문정욱	정보통신정책연구원 디지털사회전략연구실장	임 용	서울대학교 법학전문대학원 교수
박선민	구글코리아 상무	조장래	한국마이크로소프트 전무

[그림 2] 제3기 AI 윤리·신뢰성 포럼 위원단

[그림 3] 제3기 AI 윤리·신뢰성 포럼 출범식(2024.4)

1) https://eiec.kdi.re.kr/policy/materialView.do?num=249970 (KDI 경제정보센터, AI 최고위 거버넌스 'AI전략최고위협의회' 출범)
2) 이 최고위는 2024년 대통령 직속 '국가 AI 위원회'로 기능할 예정이다.
https://news.mt.co.kr/mtview.php?no=2024042513515656774 (머니투데이,2024.04.25.)

3. 맺음말

2024년에는 생성형 AI 기술의 활용 범위가 기존텍스트 생성에서 영상까지 확대되고 있다. 오픈AI의 소라(Sora), GPT-4o, 구글(Google)의 제미나이(Gemini) 1.5 등 관련 모델 활용이 늘어나며 AI의 영향력에 대한 사회적 우려가 확대되고 있다.

정부에서 '사람이 중심이 되는 『인공지능(AI) 윤리기준』'을 발표하고(2020), 이후 AI 신뢰성 기반 조성사업(2022)의 일환으로 AI 윤리·신뢰성 포럼을 시작한 지도 어느덧 3년차에 이르렀다.

지난 5월 21일 우리 정부는 영국 총리와 공동으로 'AI 서울 정상회의'를 주재했다. 회의에 참석한 정상들은 『서울 선언』과 그 부속서인 『서울 의향서』를 채택해 AI 안전·혁신·포용이라는 국제적 차원의 방향성을 제시했다. 이는 지난 2년 동안 AI 윤리·신뢰성 포럼에 참여했던 위원들이 강조해 왔던, "AI 신뢰성에 관한 국제 논의에 주도적으로 참여하자"는 제언이 이뤄진 성과로 볼 수 있겠다. 『서울 선언』과 함께 이뤄진 '서울 장관 성명'과 '서울 AI 기업 서약'에서도 AI의 책임, 발전, 혜택 등 기업의 자발적인 추구 방향을 설정했고, 국제협력 강화를 약속했다.

세계적 수준에서 AI 윤리 논의를 진행하고 있는 UNESCO도 2024년 2월 제2차 AI 윤리 글로벌 포럼을 개최했다. AI의 지속가능한 발전과 혁신을 추구하는 과정에서, '모두를 위한 AI'를 통해 공평한 혜택을 보장하는 일은 특정 행위자 단독으로는 이루기 어렵다. 정부 전략만으로 가능하지 않고, 선도적 기업들의 책임 있는 약속도 그 자체가 '안전하고 신뢰할 수 있는 AI 기술 사용'을 보장하지 않는다. 시간이 걸리더라도 AI 시스템 관련 이해관계자들이 시간을 두고 논의하는 숙의의 과정이 필요한 셈이다.

결국 국내외 AI 관련 파트너십을 강화하면서 정부 및 국제기구와의 글로벌 정합성을 고려할 수 있도록 공론장을 지속적으로 만들어야 한다. 한스 요나스는 기술 행위의 장기 효과에 대해 주목하면서, "침해 사실을 현실에서 분명하게 가시화할 수 없는 경우(미래책임에 관한 윤리의 경우)에도 인간이 이러한 일을 당하지 않도록, '보호할 수 있는 인간' 개념을 발전시켜야 한다"고 주장했다. 이러한 한스 요나스의 책임원칙이 AI 윤리를 논하는 공론장인 AI 윤리·신뢰성 포럼에서 지속적으로 논의되기를 기대한다.

※ 출처: TTA 저널 제213호