

일본인공지능학회 윤리지침에 대하여

인공지능학회윤리위원회

2017년 2월 28일

인공지능윤리위원회는 2014년 위원회를 설치한 이래 인공지능연구 혹은 인공지능기술과 사회와의 관계를 넓게 인식하고, 이를 논의하고 고찰하고, 사회에 적절하게 알려나갈 것을 추진해 왔습니다. 일본에서도 다양한 정부기관에서 인공지능과 사회에 관련된 논의가 진행되고, 또한 국제적으로도 그러한 논의가 진행되는 가운데, 인공지능학회로서도 깊은 전문지식에 기초하여 국내의 논의를 리드해 갈 역할이 있다고 생각합니다.

작년 전반기에는 이러한 취지에 기초하여 윤리위원회 내에서의 논의를 윤리강령(안)이라는 형태로 정리하였습니다. 6월의 전국대회에서 공개검토하고, 다양한 의견을 받고, 그 후 인터넷에서도 의견을 받았습니다. 이를 감안한 개정작업을 진행하고, 12월에는 개정판 윤리강령(안)을 정하였습니다. 이를 기반으로 다시 한 번 윤리전문가나 편집위원회로부터 의견도 받고, 수정을 거친 후, 이번 인공지능학회에서 ‘인공지능학회 윤리지침’이 승인되었습니다.

윤리지침의 의도는 다음과 같습니다.

윤리위원회의 목적은 인공지능기술이 초래할 긍정적·부정적 임팩트 양면에 관하여, 사회에 다양한 목소리가 있다는 것을 이해하고, 사회로부터 진솔하게 배우고, 이해에 깊이를 더하고, 사회와의 부단히 대화해 나가는 것입니다. 이 윤리지침의 의도는 앞으로의 인공지능학회와 사회와의 대화를 향한 대략적인 방침이 되는 것을 우선 신는 것입니다. 인공지능학회는 사회를 위한 연구활동을 하고 있고, 어떻게 보면 당연한 것을 당연하게 적음으로써 이를 사회에 정확하게 인식하게 하고, 대화의 기초로 하자고 생각합니다.

이러한 의도에서 이 윤리지침은 대부분이 연구자로서의 직업윤리 측면이 강하고, 당연한 것이 적혀있지만, 그 당연한 것을 정확하게 설명하는 것이 중요하다고 생각했습니다. 작년 인공지능학회 전국대회의 공개토론에서는 패널리스트가 “인공지능 연구자는 무엇을 할지 모른다고 세간에서 생각한다. 절대 미친 과학자(mad scientist)가 아니라 더 나은 사회를 위해서라고 생각하여 연구하고 있다는 것을 우선 정확히 표명해야 하며, 이러한 윤리지침을 학회 측에서 낸다는 것은 칭찬할 만한 일”이라는 내용의 코멘트도 받았는데, 윤리위원회의 의도를 정확하게 대변한 발언이라고 생각합니다.

또한 이 윤리지침은 곧 어떤 형태로든 실효성을 갖는 것을 의도하고 있는 것이 아닙니다. 예를 들어 편집프로세스로 투고논문이 윤리지침에 맞는지를 체크하는, 혹은 특정 인공지능연구가 이 윤리지침에 합치하는지를 체크하는 것 등의 체제를 구축하고자 하는 의도는 없습니다. 우선은 윤리지침을 내는 것으로 큰 틀에서의 합의를 이루고, 사회, 그리고 학회 회원들끼리 대화하면서 윤리지침이나 인공지능기술에 대하여 더 깊은 논의를 진행하고 싶다고 생각합니다. 그리고 만약 사회의 많은 사람, 혹은 많은 회원이 어떤 실효적 프로세스를 원한다면 대화를 통하여 윤리지침의 해석 및 재검토를 진행하고 싶다고 생각합니다.

그리고 이번 윤리지침에서 특징적인 것은 제9조입니다. 이는 매우 인공지능학회다운 조항이라고 생각합니다. 아시모프의 3원칙과 같은 서문참조 혹은 이 윤리지침에 따라 만들어진 인공물에 대해서도 이 윤리지침이 적용된다는 재귀성(다시 되돌아오는 성질)을 포함하고 있고, 조문으로써 흥미가 깊은 구조로 되어 있습니다. 또한 윤리위원회로서는 인공지능이 장래에 어떤 형태로 사회에서 사용될지 다양한 가능성이 있다고 생각하고 있지만, 철완 아톰이나 도라에몽이 인공지능연구에 큰 꿈을 준 일본에서는 사회 안에서 ‘구성원’으로서 인정되는 인공지능의 형태를 비교적 많은 사람이 이미지화하기 쉽고, 인류를 위한 인공지능이라는 이 윤리지침의 취지를 이해하기 쉽다고 생각합니다. 더욱이 EU에서는 로봇에 법적인격을 부여하려는 움직임이 있습니다. 게다가 제9조를 둬으로써 사람들에게 ‘사회의 구성원이란 대체 무엇인가?’ ‘인공지능이 윤리지침을 준수한다는 것은 어떤 것인가?’ 라는 다양한 의문을 던져주며, 이것이 사회 전체에서의 인공지능기술의 이해에 깊이를 더하고, 또한 인공지능의 사회 속에서의 마땅한 모습에 대한 논의가 깊어지는 것으로 연결되는 것은 아닐까 생각합니다. 이러한 논의를 만들어내는 것이 제9조의 취지입니다.

마지막으로 반복입니다만, 이 윤리지침은 인공지능연구자가 당연하게 생각하고 있는 것을 인공지능학회로서 정확하게 표명하고, 이를 통하여 연구자와 사회와의 대화에 깊이를 더하고, 사회 속에서 건전하게 인공지능기술이 이용되도록 논의를 해 나가기 위해 제정한 것입니다. 이 윤리지침을 계기로 다양한 분들의 대화의 계기가 되는 것을 기대하며, 윤리위원회는 앞으로도 활동을 계속할 것입니다.

또한 학회 회원 및 사회 전체와 대화하기 위하여 윤리지침의 공포 후 의견을 받으며, 다시 5월의 인공지능학회 전국대회의 윤리위원회공개검토회에서 논의할 예정입니다.