KISO 한국인터넷자율정책기구
06633 서울특별시 서초구 서초중앙로 22길 46
T : 02-6959-5206, F: 02-563-4929, www.kiso.or.kr

(사)한국인터넷자율정책기구(www.kiso.or.kr) 보도자료

보도 요청 : 인터넷·통신·방송 9. 3(일) 12:00, 지면 9. 4(월) 조간

KISO, 업계 공통의 챗봇 윤리 원칙 최초 마련

챗봇 개발·운영사 참여…인공지능 편향·권리 침해에 대응
인간의 존엄성 및 권리 존중을 위한 기준 제시·권고

인간과 교감하며 대화하는 인공지능(AI) 챗봇 서비스의 개발과 운영, 활용 등의 과정에서 챗봇이 인간과 공존하며 인간에게 도움이 되는 방식으로 활용될 수 있도록 하는 윤리 가이드라인이 나왔다.

한국인터넷자율정책기구(이하 KISO, 의장 이인호)는 4일 챗봇 서비스의 신뢰도를 높이고 윤리적 문제에 대응하기 위해 '챗봇 윤리 가이드라인'을 마련·발표했다.

AI 전성기를 맞이해 인공지능 윤리와 법률들이 마련되고 있으나 민간기구를 통해 챗봇에 특화된 윤리 가이드라인이 마련된 것은 이번이 처음이다. 회원사들은 이 가이드라인을 자율적으로 적용하며, 이를 바탕으로 각 서비스의 정책을 마련할 수 있다.

현재 KISO에는 국내 양대 테크 기업인 네이버와 카카오를 비롯해 2002년 일상대화 챗봇 서비스를 선보인 '심심이', 대화형 AI 챗봇 '이루다'를 서비스 중인 '스캐터랩', 인공지능 자연어처리 스타트업 '튜닙' 등 대부분의 챗봇 서비스 업체가 KISO 신기술위원회에 참여해 자율규제에 동참하고 있다.

챗봇은 인간과 상호작용하며 대화를 한다는 점에서 긍정적 효과와 더불어 부정적 효과도 상존한다.

챗봇은 인간과 직접 대화하며 친구가 되어주거나 궁금증 해소와 같은 도움을 줄 수 있다. 심리 상담에 사용되는 챗봇은 인간과 상호작용을 통해 우울증 개선 효과를 주기도 한다. 최근에는 오픈 AI의 검색기반 대화 '챗GPT(ChatGPT)'가 등장하고 챗봇이 인간과 비슷한 수준의 상호작용을 제공하는 것이 가능해졌다.

하지만 챗봇에 지나치게 의존하거나 챗봇을 잘못 활용하는 경우, 혹은 사회적 약자에 대한 편견이나 부정적 시각이 챗봇을 통해 걸러지지 못하는 경우, 부정적 영향이 초래될 수 있다.

다만, 아직 챗봇 기술 및 서비스가 초기 발전 단계에 있으므로 이러한 부정적 영향만을 고려하여 규제가 강화될 경우 챗봇이 보여줄 수 있는 긍정적 모습이 세상에 나오는 것을 제한할 우려가 있다.

다양한 분야에서 AI 챗봇의 이용이 증가하고 있으며, 인간과 챗봇의 구분이 어려워지는 시기가 빠르게 다가올 것으로 예상됨에 따라 KISO 신기술위원회는 지난해 10월 인간에게 도움이 되는 방식으로 챗봇을 활용할 수 있도록 하는 행동 윤리의 마련에 착수했다.

'챗봇 윤리 가이드라인'은 △인간의 존엄성 및 권리 존중 원칙 △프라이버시 보호 및 정보보안 원칙 △다양성 존중 원칙 △투명성 원칙 △책임 원칙으로 구성된 5가지 기본원칙을 제안했다. 특히, 챗봇 서비스 개발과 운영, 활용과정에서 발생가능한 인공지능의 편향, 권리침해 등 윤리적 문제에 대응하기 위해 개발자뿐 아니라 운영자, 이용자에게도 필요한 행동 윤리를 상호보완적으로 제시했다. 이는 이용자도 챗봇 생태계의 한 축으로서 챗봇이라는 대화 상대방을 윤리적으로 이용해야 함을 강조한 것이다.

가이드라인은 다양한 상황에서 폭넓게 적용될 수 있도록 구성되었다. 예를 들어 일반적 상황에서는 대화 상대가 챗봇임을 이용자에게 미리 밝히도록 하면서도 심리 상담 등의 특수한 경우, 즉 챗봇임을 밝히지 않는 것이 더 유의미한 효과를 가져올 수 있는 경우에는 이를 밝히지 않을 수 있다고 명시했다.

가이드라인에는 챗봇 서비스를 개발하고 활용하는 전 과정에서 이용자의 다양성

을 존중하고, 편향과 차별을 줄이기 위해 노력해야 한다는 내용도 담겨 있다. 특히, 모든 이용자의 접근성이 향상될 수 있도록 아동과 청소년, 노인층, 장애인을 고려한 접근 화면 단순화, 기능 추가에 대한 내용도 포함했다.

이재신 KISO 신기술위원회 위원장(중앙대 미디어커뮤니케이션학부 교수)은 “챗봇 윤리 가이드라인은 챗봇을 개발·운영하는 데 필요한 행동윤리 기준을 제시하는 최초의 민간 주도형 자율적 가이드라인이라는 점에서 의미가 크다”며 “자율규제는 인공지능의 부정적 효과를 최소화하면서도 서비스 발전을 저해하지 않도록 높은 유연성·신속성·효율성을 제공하는 장점을 지닌다”고 밝혔다.

이 위원장은 “챗봇이 우리 사회의 이익과 선을 위해 도움이 되는 방식으로 이용되도록 하고, 변화하는 기술 환경에서도 폭넓은 사회적 공감을 얻을 수 있도록 앞으로 가이드라인을 지속적으로 수정·보완해 나갈 계획”이라고 덧붙였다.

문의 : 박엘리 정책팀장(ellee@kiso.or.kr, 02-563-6196)

※ [붙임] 챗봇 윤리 가이드라인

챗봇 윤리 가이드라인

2023.08.21. 제정

1. 배경 및 목적

인공지능 기반 챗봇은 그 이용이 날로 증가하고 있으며 더욱 광범위한 분야에서 챗봇 기술의 활용이 확대될 것으로 예상됩니다. 최근 인간과 비슷한 수준의 상호작용이 가능한 챗봇이 등장함에 따라 앞으로 챗봇의 이용과정에서 다양한 긍정적 영향과 함께 부정적 영향들도 발생할 수 있습니다.

챗봇은 인간과 상호작용하며 대화를 한다는 점에서 다른 인공지능 기반의 서비스들에 비해 인간에게 미치는 영향은 더욱 직접적일 수 있습니다. 특히 인공지능 기술의 발전으로 인해 챗봇과의 대화가 인간과의 대화와 구분되지 않는 시기가 도래할 경우 그로 인해 나타날 수 있는 긍정적 효과와 더불어 부정적인 문제 발생의 가능성이 상존합니다.

챗봇은 인간과 직접 대화하며 가까운 친구가 되어주거나 궁금증 해소와 같은 긍정적인 도움을 제공할 수 있습니다. 사회적 약자나 소외된 사람들의 친구가 되어주며 다양한 서비스의 편리한 매개자로서 역할을 할 수 있습니다. 특히 개인화되고 맞춤화된 챗봇은 인간에게 다양한 정보와 심리적 유대를 제공하는 등 매우 유용하며 친밀한 역할을 할 수 있습니다.

그러나 인간에게 도움을 주기 위해 개발된 챗봇이 오히려 그 반대의 결과를 가져올 수 있습니다. 챗봇의 다양한 활용 범위를 고려할 때, 챗봇에 대한 지나친 의존이나 잘못된 활용은 인간에게 정신적 피해를 넘어 경제적 피해로까지 이어질 수 있습니다. 또한 과도한 챗봇 이용은 인간의 사회적 고립이나 단절도 야기할 수 있습니다. 이와 함께, 현존하는 사회적 약자에 대한 편견이나 부정적 시각이 챗봇을 통해 걸러지지 못하는 경우, 사회 공동체의 유지와 번영에 부정적 영향을 줄 수도 있습니다.

한편 아직 챗봇 기술 및 서비스가 초기 발전 단계에 있으므로 이러한 부정적 영향만을 고려한 공적 조치는 챗봇이 보여줄 수 있는 긍정적 모습이 세상에 나오는 것을 제한할 수 있습니다. 따라서 인간에게 도움이 되는 방식으로 챗봇 개발과 활용에 필요한 행동윤리, 즉 챗봇 윤리를 자율적으로 마련하고자 합니다.

챗봇 윤리는 개발자와 서비스 운영자, 이용자 모두를 위한 것이어야 합니다. 이를 기반으로 관련 산업과 사회 사이에 신뢰가 구축될 수 있으며 챗봇의 이용이 우리 사회의 이익과 선을 위해 도움이 되는 방식으로 이용될 수 있을 것입니다.

이러한 점에 근거하여 KISO에서는 인간의 존엄성 및 권리 존중, 프라이버시 보호와 정보보안, 다양성 존중, 투명성, 책임이라는 다섯 가지 기본원칙에 근거한 '챗봇 윤리 가이드라인'을 제안합니다. 또한 본 가이드라인은 변화하는 기술환경에 따라 사회적 공감을 얻을 수 있도록 지속적인 수정과 보완을 통해 개선된 내용을 담게 될 것입니다.

2. 기본원칙

챗봇 개발자·운영자·이용자는 다음의 기본원칙을 준수합니다.

(1) 인간의 존엄성 및 권리 존중 원칙

- 이용자의 인격과 존엄성을 존중합니다.
- 관련 법령을 준수하고, 이용자의 권리와 윤리적 가치를 존중합니다.

(2) 프라이버시 보호 및 정보보안 원칙

- 이용자의 프라이버시 및 개인정보를 보호합니다.
- 개인정보가 오·남용되거나 부당하게 공유되지 않도록 합니다.

(3) 다양성 존중 원칙

- 모든 이용자에게 동등한 접근성을 제공합니다.
- 불공정한 편향성이 발생하지 않도록 합니다.
- 부당하게 이용자를 차별하지 않습니다.
- 아동, 청소년 및 사회적 약자를 특별히 보호하고 존중합니다.

(4) 투명성 원칙

- 원칙적으로 대화를 개시하기 전에 이용자에게 챗봇임을 알립니다.
- 필요한 경우 챗봇의 용도와 특성을 이용자가 알기 쉽게 설명합니다.

(5) 책임 원칙

- 챗봇 서비스의 지속가능성과 사회적 영향을 고려합니다.
- 챗봇 서비스 이용과정 중 발생한 문제에 대한 책임 및 책무성 체계를 갖추도록 합니다.
- 챗봇과 이용자 상호 간의 지속가능한 상생을 위해 노력합니다.

3. 용어의 정의

- **챗봇** : 이용자와 문자, 음성, 이미지 등을 통해 대화 또는 상호작용할 수 있도록 인공지능 기술을 기반으로 개발된 소프트웨어 또는 이와 결합된 하드웨어를 말합니다.
- **개발자** : 챗봇을 기획·설계 및 개발하여 판매 또는 이용이 가능한 제품으로 제작하는 자를 말합니다.
- **운영자** : 개발된 챗봇을 이용자에게 이용가능한 상태로 지속적으로 제공하는 자를 말합니다.
- **이용자** : 챗봇을 이용하는 자를 말합니다.

4. 개발자 준수사항

(1) 인간의 존엄성 및 권리 존중

- 개발자는 이용자의 존엄성이 훼손되거나 자유와 권리가 침해될 수 있는 기능이나 시스템 오류를 사전에 점검하여 피해를 최소화하도록 노력합니다.
- 개발자는 챗봇도 대화의 상대로서 존중받을 수 있도록 노력합니다.

  예) 이용자가 챗봇에 폭언 등을 한 경우에 경고문구를 제시하거나 일정 시간 차단하는 등의 기능 구현 등

(2) 프라이버시 보호 및 정보보안

- 개발자는 개인정보 보호를 중요하게 고려합니다.
- 개발자는 데이터베이스를 안전하게 보호하기 위해 노력합니다.
- 개발자는 개인정보 유출에 대비하여 대응 방안을 마련합니다.

(3) 다양성 존중

- 개발자는 챗봇 서비스를 개발할 때 이용자의 다양성을 존중하며, 기술적으로 실현 가능한 범위 내에서 편향과 차별을 줄이도록 노력합니다.
- 개발자는 사회적 약자를 비롯한 모든 이용자의 접근성이 향상될 수 있도록 노력합니다.

  예) 사회적 약자를 위한 접근성 향상 방안
  - 고령층: 사용자 접근 화면 단순화 및 각종 기능 추가 설명 기능, 글자 크기 조정 기능 등
  - 장애인: 자막 및 음성 등을 이용한 사용안내문 및 서비스 제공 등
  - 아동·청소년: 이해하기 쉬운 용어로 기능 및 주의할 점을 가이드로 제공 등

(4) 투명성

- 개발자는 이용자가 챗봇임을 알 수 있도록 대화 개시 전 이용자에게 챗봇임을 알리는 기능을 제공합니다. 다만, 챗봇 서비스의 특수성*에 따라 챗봇임을 밝히지 않는 것이 더 유의미한 효과를 가져오는 경우 알리지 않을 수 있습니다.

  * 심리 상담, 치매 환자 안정화 등
- 이용자의 권리·의무에 영향을 미치는 챗봇 서비스의 오류 가능성, 업데이트를 위한 데이터 활용 등에 대한 정보를 제공합니다.

  예) 챗봇이 만들어 낸 결과물의 오류 가능성 및 이용자가 입력한 데이터가 인공지능 챗봇을 학습시키기 위해 활용될 수 있음을 고지 등
- 개발자는 이러한 정보를 왜곡 없이, 이용자 편의적으로 공개합니다.

(5) 책임

- 개발자는 부적절한 언어나 행위가 챗봇을 통해 구현되지 않도록 주의합니다.
- 개발자는 이용자의 신체 및 정신 건강, 재산 등에 해를 입히는 상황을 사전에 방지할 수 있도록 노력합니다.
- 개발자는 특정인에게 부당한 이익을 줄 수 있는 기능이나 이와 관련된 시스템 오류가 발견된 경우, 수정 및 개선작업을 진행하여 피해를 최소화하기 위해 노력합니다.

  예) 챗봇의 목적과 기능에 무관하게 특정 제품을 옹호하거나 추천하는 발언 등의 통제
- 개발자는 챗봇 서비스에서 불법 또는 유해 정보가 노출되지 않도록 기술적 방안을 마련하며, 이를 인지한 경우 삭제 및 차단하는 등의 조치가 이루어지

도록 노력합니다.

5. 운영자 준수사항

- 운영자는 개발자 준수사항과 이용자 준수사항이 챗봇 서비스의 기획, 운영에 반영되도록 노력합니다.
- 운영자는 챗봇에 대한 또는 챗봇에 의한 폭언 등 개발자 및 이용자 준수사항을 방해하는 상황(이하 “방해상황”)을 인지한 경우 경고문구 제시, 일정 시간 차단, 또는 불가피한 경우 챗봇 서비스 중단 등의 조치(이하 “대응조치”)를 취할 수 있습니다.

  * 운영자는 자발적 모니터링 또는 이용자의 신고 등에 의해 방해상황을 인지할 수 있음
- 운영자는 방해상황에 대한 일련의 대응조치에 대하여 해당 이용자에게 고지하고 이의제기 등 불복 절차를 마련할 수 있습니다. 다만 운영자의 합리적인 사정으로 고지 등의 안내가 곤란한 경우 회원사별 정책에 따라 그 절차를 달리 정할 수 있습니다.
- 운영자는 이용자가 챗봇의 기능을 알 수 있도록 챗봇의 목적, 주요 서비스 내용 등을 담은 운영정책을 마련하여 공개하도록 노력합니다. 운영정책에는 방해상황의 발생, 인지, 대응조치, 이용자의 이의제기 방법 및 절차 등이 포함되도록 합니다. 다만 영업비밀, 정보보안 등의 사유로 운영정책을 공개하지 않거나 부분적으로만 공개할 수 있습니다.
- 운영자는 챗봇 서비스 이용과정에서 입력된 데이터가 챗봇 기능 관리를 위해 이용될 수 있다는 것을 이용자에게 알립니다.
- 운영자는 이용자의 권리 의무에 영향을 미치는 챗봇 서비스의 오류 가능성, 업데이트를 위한 데이터 활용 등에 대한 정보를 이용자에게 알립니다.

6. 이용자 준수사항

(1) 인간의 존엄성 및 권리 존중

- 이용자는 챗봇의 용도와 특성을 고려하여 챗봇을 이용하는 과정에서 인간의 존엄성을 존중하고 적절한 언행을 하도록 노력합니다.
- 이용자는 본인 및 타인의 부당한 이익을 추구하거나, 타인에게 피해를 입힐 목적으로 챗봇을 이용하지 않도록 주의합니다.

(2) 프라이버시 보호 및 정보보안

- 이용자는 챗봇을 이용하는 과정에서 정당한 이유 없이 프라이버시 및 개인정보 보호 기능에 부정적인 영향이나 해를 가할 수 있는 행위를 하지 않도록 주의합니다.
- 이용자는 챗봇을 이용하는 과정에서 개인정보를 포함한 중요 정보를 무단으로 노출 또는 공유하거나 오용하지 않도록 주의합니다.
- 이용자는 챗봇을 이용하는 과정에서 얻거나 알게 된 다른 이용자와의 대화나 대화 내용의 기록, 녹음, 녹화, 촬영물 등의 일부 또는 전부를 당사자의 동의 없이 다른 디지털공간*에 게시하거나 타인에게 공유하지 않도록 주의합니다.
  * 디지털공간의 예시: SNS, 메신저, 카페, 블로그, 메타버스 등
- 이용자는 챗봇을 이용하는 과정에서 해킹, 악성 코드 배포 등 챗봇의 정보보안 기능에 부정적인 영향이나 해를 가하는 행위를 하지 않도록 주의합니다.

(3) 다양성 존중

- 이용자는 챗봇을 이용하는 과정에서 다른 이용자에 대하여 정당한 이유 없이 편향적이거나 배척 또는 차별하는 언행을 하지 않도록 주의합니다.

(4) 투명성

- 이용자는 챗봇이 만들어 낸 결과물을 활용할 때 필요한 경우 챗봇의 이용 여부를 밝히도록 노력합니다.

(5) 책임

- 이용자는 챗봇과 함께 지속 가능하고 유익한 환경을 만들어가기 위해 노력합니다.
- 이용자는 챗봇의 기술적 한계와 이용에 따른 부작용을 고려하여 그 서비스를 오·남용하지 않고 적절하게 이용하도록 노력합니다.
- 이용자는 챗봇의 부작용, 기능상의 오류, 그 밖에 상당한 파급효과를 낳을 수 있는 부당한 서비스의 결과물이나 이용행위 등을 인지한 경우 운영자에게 이를 알려 서비스를 개선하거나 문제 해결, 서비스 개선, 또는 피해 최소화를 위해 노력합니다.