

한국고용정보원, 국민이 안심할 수 있는 ‘인공지능(AI) 고용서비스 윤리원칙’ 제정

사람 중심·신뢰·공정·사회적 가치 4대 핵심 가치 마련
실행력 강화를 위한 가이드라인 제정 추진

한국고용정보원(원장 이창수)은 인공지능(AI)이 고용서비스 전반에 확대되는 상황에서, 국민이 더욱 안전하고 신뢰할 수 있는 서비스를 이용할 수 있도록 ‘인공지능(AI) 고용서비스 윤리원칙’을 제정했다고 10일 밝혔다.

이번 윤리 원칙은 정부(과학기술정보통신부)의 ‘인공지능 윤리기준(2020)’을 기반으로 한국고용정보원의 고유업무와 현장 요구를 반영해 마련된 것으로 ▲사람 중심 ▲사회적 가치 실현 ▲신뢰 ▲공정 등 4대 핵심 가치를 중심으로 구성되어 있다.

□ 4대 핵심 가치 및 10대 윤리 원칙

① (사람 중심) 인공지능(AI)이 사람의 결정을 대신하는 것이 아니라 이용자의 선택권·존엄성을 우선 고려하도록 설계해, 누구나 안전하게 서비스를 이용할 수 있도록 한다.

* ①이용자 권리 보장 ②이용자 개인정보보호 ③권익 침해방지

② (사회적 가치 실현) 공공기관으로서 인공지능(AI) 기술이 사회 전체의 고용 기회 확대와 고용 생태계 발전에 기여하도록 운영한다.

* ④사회적 책무 이행 ⑤고용생태계 연계성

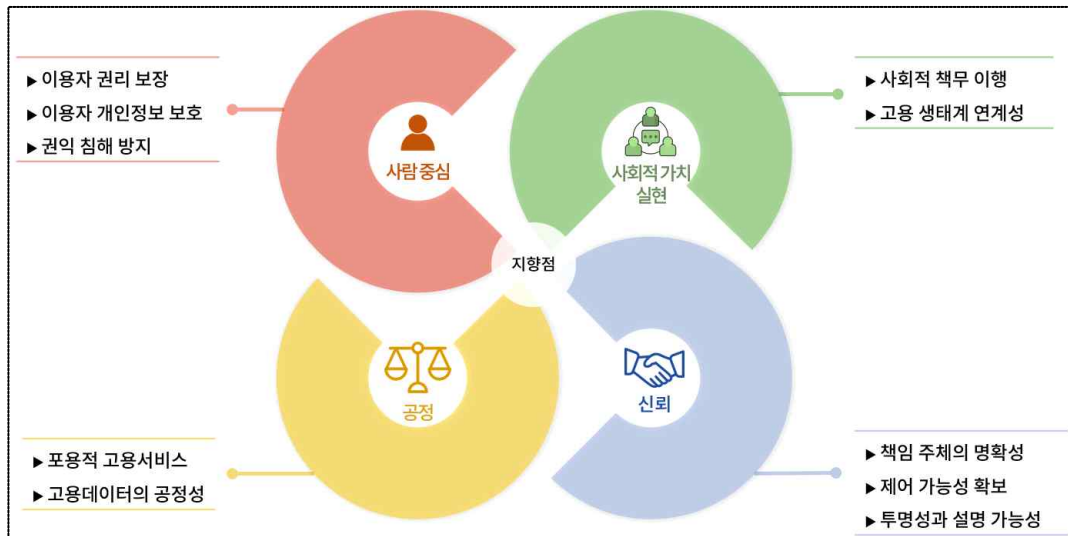
③ (신뢰) 인공지능(AI)이 어떤 근거로 판단했는지를 국민이 이해할 수 있도록 설명 가능성과 투명성을 강화하며 결과에 대한 책임성을 높인다.

* ⑥책임 주체의 명확성 ⑦제어 가능성 확보 ⑧투명성과 설명 가능성

- ④ (공정) 데이터 편향으로 특정 집단이 불이익을 받지 않도록 지속적으로 점검해 모든 국민에게 차별 없는 고용서비스를 제공한다.

* ⑨포용적 고용서비스 ⑩고용데이터의 공정성

[그림] 한국고용정보원 AI 고용서비스 윤리 원칙



한국고용정보원은 윤리원칙이 실제 서비스 운영에 적용될 수 있도록 ‘인공지능(AI) 기반 고용서비스 운영 가이드라인’을 새롭게 제정할 계획이다.

이 가이드라인은 인공지능(AI) 설계·운영·평가 전 과정에서 준수해야 할 절차를 세부적으로 제시해 윤리원칙의 실행력을 높이는 역할을 하게 된다.

또한, 정부가 발표하는 최신의 인공지능(AI) 윤리 정책과 관련 기준을 즉시 반영할 수 있는 체계를 구축해 책임 있는 인공지능(AI) 운영을 한층 강화할 방침이다.

한국고용정보원 이창수 원장은 “이번 윤리 원칙은 책임 있는 인공지능(AI) 활용을 위한 중요한 기준”이라며 “앞으로도 신뢰할 수 있는 인공지능(AI) 고용서비스를 통해 국민이 안전하고 편리하게 고용서비스를 이용하도록 최선을 다하겠다.”라고 밝혔다.

담당 부서	국가고용정보망운영본부 AI 고용서비스팀	책임자	팀 장	조인성 (043-870-8450)
		담당자	대 리	정재훈 (043-870-8765)

참고

한국고용정보원 AI 고용서비스 10대 윤리 원칙

원칙	정의
①이용자 권리 보장	인공지능의 기획, 개발, 활용, 운영 전반에 걸쳐 이용자의 인권을 최우선으로 고려하고 보호하는 것 을 의미함. 여기서 인권이란 인간의 존엄성과 자유, 평등, 자기 결정권 등 보편적이고 기본적인 권리를 포함함.
②이용자 개인정보 보호	인공지능 시스템의 기획, 개발, 운영, 평가 과정 전반에 걸쳐 이용자의 개인정보 및 사생활을 보호 해야 함. 데이터의 수집·이용·저장·처리 과정에서 이용자의 프라이버시에 대한 최소한의 개입과 이용자가 자신의 정보를 통제할 수 있는 환경을 보장하는 것을 의미함.
③권익 침해 방지	인공지능이 인간에게 해를 끼치거나 권리를 침해하지 않도록 개발 및 활용 과정에서 발생 가능한 위험 요인을 사전에 식별·차단하고 기술의 오남용을 방지해야 함.
④사회적 책무 이행	공공기관의 인공지능 개발 및 활용이 이용자 전체의 이익을 지향하며, 사회 전체에 긍정적 영향을 미치는 방향으로 이루어지도록 책임을 다하는 것을 의미함.
⑤고용 생태계 연계성	인공지능은 사회 각 분야의 다양한 이해관계자들과의 협력과 연계 를 통해 개발·활용되어야 함. 지속 가능한 생태계 구축을 위해 민간, 공공, 학계, 시민사회 등의 연대가 요구되며, 기술이 공동체적 가치와 연결성을 높이는 수단으로 작용해야 함.
⑥포용적 고용서비스	인공지능 시스템이 특정 개인이나 집단에 불리하게 작용하지 않도록 다양한 사회적 배경, 특성, 조건을 존중 하며 설계·운영되어야 함. 인종, 성별, 연령, 지역, 장애 등 다양한 요소가 적절히 고려되도록 데이터 구성 및 알고리즘 설계 전반에서 형평성과 포용성을 확보 하는 것을 목표로 함.
⑦고용데이터의 공정성	데이터 수집과 활용의 전 과정에서 편향성을 최소화 하고 데이터 품질과 위험을 관리하는 것을 의미함. 이는 데이터의 정확성, 대표성, 최신성을 확보하고 데이터에 내재한 편향을 식별·보정하여 인공지능이 공정한 의사결정 을 수행하도록 하는 것을 포함함.
⑧책임 주체의 명확성	공공 고용서비스의 신뢰성을 보장하기 위해 인공지능 활용 과정에서 발생하는 영향에 대해 제도적 책임 주체를 명확히 해야 함. 인공지능 개발 및 활용의 전 과정에서 발생할 수 있는 영향에 대해 명확한 책임 주체 를 지정하고, 그에 따른 책임과 조치 를 사전에 정의해야 함.
⑨제어 가능성 확보	인공지능의 개발부터 활용에 이르기까지 전 과정에서 시스템이 예기치 못한 위험을 유발하지 않도록 설계·관리하며, 명백한 오류나 침해 발생 시 이를 추적·기록 하고 관리자가 인지 및 제어할 수 있는 장치를 마련해야 함.
⑩투명성과 설명 가능성	고용서비스가 이용자에게 미치는 영향을 고려하여 인공지능 활용 여부와 그 과정에서 발생 가능한 위험요소를 사전에 충분히 고지 하고, 이해 가능한 방식으로 설명하려는 노력이 요구됨.