KOREAN
WOMEN'S
DEVELOPMENT
INSTITUTE

젠더리뷰

기획특집

인공지능(AI) 기술 발전과 성인지적 대응 전략

KOREAN WOMEN'S DEVELOPMENT INSTITUTE

공정성을 넘어: AI 채용도구 형평성 제고전략

이지은 연세대 문화인류학과 부교수
신민정 연세대 문화인류학과 박사과정
신지연 연세대 문화인류학과 박사과정

1. 들어가며

AI 채용도구는 구인 광고와 구직자 매칭, 이력서 및 자기소개서 스크리닝, 인적성 검사와 유사한 역량검사와 면접 등 채용 전 단계에서 활용되는 AI 활용 어플리케이션을 가리킨다. 이 글에서는 흔히 'AI 면접'이라고도 불리는, 역량검사 및 면접 단계에서의 AI 채용도구에 초점을 맞추고자 한다. AI 채용도구는 채용과정 관리에 소요되는 비용과 시간을 절감하기 위해 도입되었으며 코로나19로 인해 대면 실시가 어려웠던 지필 인적성검사 등을 대체하면서 급격하게 확산되었다. 채용비리 등 불공정 채용 문제로 여러 차례 사회적 논란이 되어 왔던 공공기관에서는 AI 채용도구를 사용하여 채용의 투명성, 공정성, 효율성을 제고할 수 있을 것으로 보았다. 실제로 2018년 AI 역량검사를 인사 프로세스에 도입한 한국자산관리공사는 우수한 인사혁신사례로 소개되기도 했다(인사혁신처, 2018).

이제 구직자들에게 AI 면접은 그 자체로 낯선 것은 아니다. 하지만 그것이 어떤 방식으로 작동하는 것인지, AI 채용도구가 내리는 결정은 과연 '공정'한지, 그 결과가 실제 채용에 어떤 영향을 미치는지 등은 여전히 의문으로 남아 있다. 이 불확실성은 많은 불안을 야기한다. 예를 들어 구직자는 컴퓨터 앞에서 정확히 무슨 능력을 측정하는지 모를 게임을 하고 이것이 직무와 어떻게 연계가 되는지 알 수 없다고 느낀다. 또 면접관이 아닌 모니터를 앞에 두고 자기를 소개하는 방식은 구직자들을 불안하게 한다. 이러한 불안을 겨냥하여 'AI 면접코칭'이라는 새로운 시장이 등장하기도 했다. AI 채용도구가 어떤 방식으로 작동하는지, 무엇을 어떤 방식으로 측정하며 얼마나 신뢰할 만한 것인지는 그 도구를 도입해 활용하는 기업의 인사담당자들에게도 명확하지 않다. 각각의 도구에 대한 이해 수준 역시 인사담당자에 따라 상이한 상황이다. 현재 국내에서 사용하고 있는 채용도구들은 소수의 민간 기업에 의해 개발된

것들이다. 여기에서 '블랙박스'와도 같은 AI 채용도구의 불투명성이라는 문제가 발생한다. 이런 상황은 AI 채용도구의 작동원리와 학습 데이터의 수집 및 가공 방식 등과 관련해서도 불안을 야기한다. 인사담당자들조차 각 업체가 자사에서 개발한 도구의 우수성과 신뢰성을 홍보하기 위해 공개하는 내용 이상을 파악하기 어려운 상황이다.

이러한 문제를 보완하기 위해 정보통신정책연구원(KISDI)과 한국정보통신협회(TTA) 등이 각각 윤리점검표 개발과 AI 신뢰성 검증 작업을 수행하고 있다. KISDI의 인공지능 윤리기준 실천을 위한 자율점검표(과학기술정보통신부·정보통신정책연구원, 2023)는 다양성 존중 및 침해금지 등의 조항을, TTA의 신뢰성 검증(과학기술정보통신부·한국정보통신기술협회, 2024)은 수집 및 학습된 데이터의 편향 및 인공지능 모델의 편향 제거, 인공지능 모델 명세 및 추론결과에 대한 설명 제공, 인공지능 시스템의 설명에 대한 사용자의 이해도 제고 등의 항목을 포함하고 있다. 형평성 제고라는 관점에서 볼 때 이러한 노력들에서 가장 주목할 만한 부분은 '편향(bias)'의 최소화에 대한 강조이다. 〈신뢰할 수 있는 인공지능 개발안내서〉와 〈인공지능윤리기준 자율점검표〉에서는 각각 인공지능의 신뢰성을 높이고('신뢰') 다양성을 보장하고 차별을 방지하기 위해('윤리') 성별, 연령, 인종 등을 포함해 차별적 결과를 낼 수 있는 인구학적 속성과 관련한 편향들을 찾아내고 이러한 속성들의 영향력을 완화시켜야 한다는 것이다. 데이터의 라벨링 과정에서의 편향을 방지하기 위해 다양한 배경의 작업자를 확보하고 라벨링과 관련한 가이드라인을 만들 것 역시 권고되고 있다. AI의 젠더 편향에 대한 문제 제기가 지속적으로 이루어졌던 것을 감안할 때 편향 제거를 위한 여러 노력들은 의미 있는 움직임이다.

그럼에도 불구하고 이러한 노력들은 AI 채용도구의 형평성을 담보하기에는 부족함이 있다. 형평성(equity) 제고가 개개인의 차이와 특수한 조건들, 구조적 차별이나 주변화의 결과로 발생할 수 있는 격차를 최소화하는 것을 요구한다고 볼 때, 정량적 수준에서 데이터와 모델의 편향을 최소화하는 기술적 방식의 공정성(fairness) 확보는 이를 위한 충분조건이 될 수 없기 때문이다. AI 채용도구의 내적 타당성 수준을 넘어 그것이 실제 현장에 도입되어 활용되는 과정에서 어떤 효과를 만들어내는지에 대한 검토 역시 필요하다. 기존 채용과정에서의 차별적 전제나 관행이 AI 채용도구에서 여전히 반복되는 것은 아닌지, 보다 다양한 지원자들을 포용하는 데 AI 채용도구가 장애물이 될 가능성은 없는지 등 역시 중요한 쟁점이다.

AI 채용도구 그 자체의 '공정성' 확보를 넘어 그것이 활용되는 실제 채용과정에서 형평성을 제고하기 위한 실천적 전략으로, 본고에서는 AI 채용도구 활용의 효과를 파악하기 위한 장치 마련, AI 채용도구의 개발 및 설계 과정에 대한 면밀한 검토, AI 채용도구의 사용자 인터페이스 개선 등을 제안한다. 이에 앞서 국내 주요 AI 채용도구들의 특징 및 각 업체의 공정성 확보전략을 살펴보고, AI 채용도구 형평성 제고전략과 관련해 구체적인 쟁점과 제안사항을 제시하고자 한다.

2. 국내 AI 채용도구 활용 현황

국내의 AI 채용도구 시장에서 가장 널리 활용되고 있는 A사의 채용도구는 게임을 기반으로 한 역량

검사가 포함되어 있다는 특징을 가지고 있다. 이 채용도구는 지원자가 자신과 관련된 항목들에 대한 응답을 분석하는 자기보고식 검사와 게임을 기반으로 한 과제 수행 중 드러나는 행동 및 반응 특성을 측정하는 역량검사, 자동화된 단방향의 인터뷰 영상에서 나타나는 시각 및 음성적 관찰특성을 분석하는 영상면접의 세 파트로 이루어진다.

A사는 자사 웹사이트에 공개한 백서를 통해 AI 채용도구의 개발 원리와 신뢰성 검증 자료, 공정성을 확보하기 위한 장치에 대한 정보를 제공하고 있다. 백서는 AI 역량검사 평가결과의 일관성, 역량검사 성적과 재직자 성과 사이의 상관관계 등에 대한 통계적 검증을 통해 타당성을 입증했다고 밝히고 있다. A사는 기존의 응시자 데이터를 분석한 결과 남성과 여성의 통과율 차이가 4/5법칙(취약집단 통과율이 상대집단의 80%일 때 차별로 간주하는 미국고용평등기회위원회 기준)에 미치지 않아 통계적으로 성별에 따른 차별적 효과가 없다고 판단했다고 설명한다.

개발과정에서의 공정성 확보 장치로는 첫째, 영상면접 분석 모델의 데이터셋을 구축하는 과정에서 성별 비율의 균등한 유지가 있고 둘째, 데이터 라벨링 과정에서 편견 개입 방지를 위한 차별 유발 요소(출신학교, 지역 등) 배제 등이 있다. 백서에는 포함되지 않았지만, 게임 과제를 테스트한 결과 특정 집단이 다른 집단에 비해 성적이 현저히 좋지 않은 경우에는 그 과제를 폐기하는 등의 방식으로 역량검사의 공정성을 기하고 있다고 밝히기도 했다.

A사에 비해 뒤늦게 채용도구 시장에 진입한 B사의 AI 채용도구는 실제 면접관들의 "인사이트(insight)"를 학습한 AI 솔루션을 표방하고 있다. 평가는 실제 지원자들의 면접 영상으로부터 비언어적 요소들을 평가하는 소프트 스킬 평가와 특정한 경험에 대한 응시자의 응답으로부터 역량을 평가하는 행동사건면접(BEI: Behavior Event Interview) 평가로 구성되어 있다. 평가에는 전문 면접관들이 라벨링한 비대면 면접 응답 영상 및 텍스트 데이터를 딥러닝으로 학습한 AI가 활용된다. B사는 채용업체가 원하는 경우에 한해, 채용업체에서 자체적으로 선정한 면접관이 직접 라벨링을 수행하여 각 업체에 적합한 모델을 만들 수 있도록 한다.

B사는 TTA의 AI 신뢰성 검증 기준을 통과하였다는 점, KISDI와 함께 인공지능 윤리점검표를 개발하였다는 점 등 외부기관과의 검증 및 협업을 통해 신뢰성 및 윤리성을 확보하려 노력했다는 점을 강조한다. B사 웹사이트의 '신뢰성'과 관련된 페이지에는 TTA의 신뢰성 검증 기준 항목과 함께 AI 면접관과 라벨러인 (인간) 전문면접관의 평가 데이터 간 비교검증, 평가자 간 신뢰도 검증, 평가 목표와 평가결과 사이의 관련성 등에 대한 통계적 검증결과가 게시되어 있다.

'윤리성'과 관련해 B사는 KISDI와의 협업으로 인공지능 윤리 자율점검표를 개발하고 준수하고 있음을 강조한다. 이 윤리점검표의 형평성 관련 항목으로는 인권보장(사용자의 특성에 근거한 차별 여부), 다양성 존중(취약계층의 접근 가능성, 데이터의 편향 최소화를 위한 절차 마련, 다양한 사회, 경제적 배경을 가진 라벨러의 참여, 언어습관이나 표현방식, 시선 처리 등에 근거한 차별 방지) 등이 있다. 다만 데이터셋이나 데이터 라벨러(면접관 구성)의 다양성 확보를 위한 전략들은 구체적으로 기술되어 있지 않다.

이외에 AI를 활용한 자기소개서 분석 서비스로 출발한 C사는 최근 AI를 이용한 면접 서비스로 사

업을 확장하고자 시도하고 있다. 하지만 이 업체의 AI 면접 서비스는 아직 널리 이용되고 있지는 않은 것으로 보인다. C사의 AI 채용도구는 대화형 AI 면접 외에 자기소개서를 통해 드러나는 역량 평가 및 면접질문 생성, 기준 미달 자기소개서 선별 등 자기소개서를 분석하는 서비스와 심리학을 기반으로 한 역량검사 등을 포함하고 있다. 특히 자기보고식의 심리학을 기반으로 한 역량검사를 '부적응 가능성 예측 특화 역량검사'로 홍보하고 있다는 점이 특기할 만한 부분이다.

기업 웹사이트 등을 통해 신뢰성이나 공정성, AI 윤리와 관련된 정보를 제공하고 있는 앞의 두 업체와 달리 C사는 이와 관련한 구체적인 정보를 제공하고 있지 않다. 다만 C사는 채용 담당자들을 독자로 설정하고 있는 블로그를 운영하고 있으며 이 블로그에 게시된 포스트에서 편향성 문제 해결을 위한 방안을 간단히 설명하고 있다. 여기서는 학습 데이터에서 편견을 유발할 수 있는 개인정보를 마스킹 처리한 상태로 AI 채용도구를 개발하여 활용하고 있다고 언급하고 있다.

3. AI 채용도구에 대한 감사(audit)의 필요성

AI 채용도구 개발업체들은 개발과정에서 편향의 가능성을 최소화하기 위해 여러 장치들을 마련했다고 밝히고 있다. 하지만 성별이나 그외의 특성에 의한 차별이 실제로 발생하는지에 대해서는 충분한 검증이 이루어지지는 않은 상태이다. 업체들이 공개하고 있는 검증결과에서 초점이 되는 것은 AI 채용도구 자체의 신뢰성과 타당성이다. 고객사인 채용업체가 도구를 도입할 때 그 신뢰성과 타당성이 중요한 고려사항이라는 것을 감안한다면 AI 채용도구의 내적 타당성에 대한 개발업체의 자체검증은 개발업체의 이해관계로부터 자유롭다고 보기 어렵다. 한편 도구 자체의 내적 타당성에 대한 검증만으로는 실제 채용과정에서 발생할 수도 있는 차별 가능성을 파악하기 어렵다는 문제도 있다. AI 채용도구가 업데이트되는 경우, 혹은 측정역량의 선정 및 가중치 부여나 고객사 상황을 고려한 자체 데이터 라벨링 등 커스터마이징이 이루어지는 경우 등 다양한 경우를 상정할 수 있다. 이럴 때 AI 채용도구가 실제 활용되는 현장에서 지원자 평가에 어떤 영향을 미치는지에 대한 지속적인 조사가 필요하다고 할 수 있다.

이 지점에서 AI 채용도구에 대한 편향감사(bias audit)를 의무화한 뉴욕시의 NYC144법을 참고해 볼 만하다. 이 법은 자동화된 의사결정 기반의 채용도구가 고용상의 결정에서 중요하게 활용되기 위해서는 해당 도구가 최근 1년 이내에 편향감사를 받은 이력이 있어야 한다고 규정하고 있다. 또한 이를 도입하는 채용업체는 감사결과의 요약이나 배포일자 등을 웹사이트에 공개할 것을 의무화하고 있다. 편향감사는 개발업체가 아닌 독립된 감사인에 의해 이루어지며, 개발업체는 이 도구를 활용하여 진행되었던 실제 채용과정에서 수집된 데이터(historical data)를 감사인에게 제공해야 한다. 감사인은 이렇게 제공된 데이터로부터 성별, 인종/종족(race/ethnicity), 그리고 성별과 인종/종족의 교차범주에 대한 채용도구의 선택률(selection rate) 및 영향비(impact raio)를 계산함으로써 채용도구가 차별적인 결과를 발생시켰는지 확인한다. 실제 채용과 관련된 데이터가 통계적으로 유의미한 편향감사를 하기에 부족한 경우에 한해 테스트 데이터를 활용할

것이 허용되는데, 이 경우에는 그 사유를 감사결과 보고서에 포함하도록 한다.

NYC144법은 실제 채용과정에서 생산된 데이터를 분석함으로써 이 도구의 내적 타당성을 넘어 그것이 실제 채용과정에 미치는 영향을 파악하고 이를 지원자들에게 고지할 것을 의무화한다. 이해당사자가 아닌 제3자가 감사를 담당하고 그 결과를 고지하게 함으로써 검증 절차의 투명성을 제고할 수 있게 된다. 또한 각 채용업체에 AI 채용도구의 감사 이력을 고지할 것 등을 포함한 설명의무를 부과함으로써 채용업체들이 도구의 도입과 활용에서 보다 책임성을 가지게 되리라고 기대할 수 있다. 또 1년 단위로 이루어지는 편향감사의 의무화는 시간의 경과와 도구의 변화에 따른 영향을 지속적으로 추적할 수 있게 한다. 마지막으로 성별과 인종/종족 등 주요한 두 변수뿐 아니라 그 교차범주 역시 고려함으로써 단일 변수에 초점을 맞출 때 발생할 수 있는 문제들을 방지하는 효과를 기대할 수 있다.

국내에 이와 같은 편향 감사를 도입하는 데는 중요한 난점이 있다. 입사지원서에 성별을 포함하여 차별과 관련한 인구학적 데이터들을 기입하지 않는 것이 권고되어 온 상황에서 AI 채용도구 활용의 영향을 파악하기 위한 데이터 수집 자체가 이루어지기 어렵다는 것이다. 공공기관을 중심으로 이루어져 온 블라인드채용은 AI 채용에서 성별과 관련하여 영향이 발생했는지 여부조차 파악할 수 없다는 '알리바이'가 되기도 한다.

여기서 미국 고용평등기회위원회(EEOC)의 '지원자의 인구학적 정보' 설문(OMB No: 3046-0046)을 참고해 볼 수 있을 것이다. EEOC는 응시서류 접수 시 별도로 정보수집에 자발적으로 동의한 지원자들에 한해 성별, 인종, 종족, 장애 및 질병 정보 등에 관한 설문에 응답하도록 하고 있다. 이를 참고하여 차별을 발생시킬 수 있는 인구학적 데이터들을 익명으로 수집하고, 채용담당자들이 직접 접근할 수 없는 별도의 디지털 데이터베이스로 만들어 관리하는 방식을 고려해 볼 필요가 있다. 이러한 별도 설문을 도입할 경우 인구학적 데이터의 수집 및 보관에 관한 구체적인 의무사항을 설정해야 한다. 또한 구직자들의 취약한 위치와 불안을 감안하여 정책적 목적에 대한 충분한 홍보와 함께 미응답시에는 불이익이 없다는 것 역시 명확히 고지해야 한다.

이와 더불어 편향을 감사하기 위해 정부 차원에서 벤치마크 데이터셋을 구축하여 활용하는 방안을 검토해 볼 수 있다. 특히 모든 AI 채용도구가 채택하고 있는 음성 및 영상 분석을 통한 소프트스킬 평가 과정에서 차별적 효과가 발생하는지를 확인하는데 활용될 수 있다. 성별, 연령, 인종, 출신지역, 장애여부 등 다양한 변수들을 폭넓게 고려한 데이터셋을 제작한다면, AI 채용도구 개발과정에서 충분히 고려되지 않는 차별요인들을 검토할 수 있을 것이다.

4. 채용도구 개발 및 설계과정에 대한 이해도 제고

형평성 관점에서 채용도구를 검토하기 위해서는 타당성, 신뢰성, 공정성을 기술적으로 평가하는 정량적 감사를 넘어, 그 도구가 기반하는 전제들이 차별적이지 않은지에 대한 정성적 평가가 필요하다. AI 채용도구의 감사는 여전히 차별을 성별 등 특정 범주에 속한 인구집단과 관련한 통계적 편향의 문제로 다루고 있다는 점에서 제한적이다. 개별 지원자들 사이의 차이, 구조적 차별의 효과, 채용 전반을

관장하는 전제들에 대한 고려 없이는 채용과정에서의 차별과 형평성의 문제를 효과적으로 다룰 수 없기 때문이다. AI 채용도구의 '사회-기술적 매트릭스(socio-techincal matrix)'는 AI 채용도구의 정성적 평가를 위한 유용한 도구이다(Sloane et al., 2022). 여기에는 AI 채용도구의 기술적 특징뿐 아니라 그것이 구축되는 사회맥락 안에서 평가할 수 있도록 하는 일련의 항목이 포함되어 있으며, 이와 관련한 정보를 얻기 위해 어떤 방법들을 사용하는지가 제시되어 있다. 조사 항목들로는 채용의 어떤 단계에서 AI 채용도구가 활용되는지(단계, funnel stage), 그 활용목적은 구체적으로 무엇인지(목적, goal), 각각의 도구는 어떤 데이터를 활용하여 학습하고 테스트하고 작동되는지(데이터, data), 그 도구가 어떤 기능을 하며 무엇을 위해 최적화되어 있는지(기능, function), 그 도구가 왜 유용하다고 여겨지는지, 지원자 데이터와 채용담당자의 목표 사이에 어떤 관계가 있다고 전제되며 그것이 채용에 있어 어떤 영향을 미칠 수 있을지(전제, assumption), 그리고 각각의 AI 채용도구가 기반하고 있는 평가방식 등이 근거하고 있는 이론 등이 차별적인 성격을 가지고 있지는 않은지(인식론적 근간, epistemological roots) 등이 있다. 이 항목들에 대해 조사하는 과정에서 조사자는 AI 채용도구에 대해 매우 구체적인 지식을 얻을 수 있으며 각각의 도구가 내포하고 있는 문제들을 면밀하게 검토할 수 있다.

이 평가도구의 유용성을 이해하기 위해 해당 도구를 고안한 연구진의 파이메트릭스(Pymetrics)와 휴먼틱(Humantic)의 분석 사례를 간단히 살펴보겠다. 파이메트릭스는 지원자들이 수행한 게임 성적(데이터)을 토대로 그 사람의 능력(ability)을 측정하고(목적), 게임수행능력을 지원자 스크리닝에 활용한다(기능). 이 평가도구는 게임을 통해 드러나는 능력이 직장에서의 성공 여부를 예측가능하게 한다는 전제를 기반으로 하고 있으며, 여기에는 지능을 포함한 능력이 선천적인 것이라는 관점이 반영되어 있다(인식론적 근간). 한편 이력서, 링크드인, 트위터 프로필 정보 등을 분석하여(데이터) 지원자의 성격을 파악하고(목적) 이를 업무와 매칭에 활용하는(기능) 휴먼틱이라는 도구는 성격이 업무 적합성을 적절하게 예측할 수 있는 지표라는 전제를 기반으로 하고 있으며, 이는 다시금 성격 유형에 대한 이론들, 즉 성격은 시간이 경과하더라도 안정적으로 유지되며 성과를 예측할 수 있게 해준다는 관념에 기반하고(인식론적 근간) 있다(Sloane et al., 2022).

이처럼 기술사회적 매트릭스를 활용한 분석은 AI 채용도구를 위해 달성하고자 하는 목표가 무엇인지에 더해, 그것이 어떤 사회문화적 편견이나 차별적인 전제에 기반하고 있지 않은지 점검할 수 있게 해준다. 능력이나 성격 등은 변화하지 않는 선천적인 특질이 아니며, 그것을 정의하고 측정하는 방식에 따라 가변적일 수 있다는 점을 고려한다면 앞의 AI 채용도구의 형평성에 대한 비판적인 질문이 제기될 수 있다. 능력이나 성격을 정의, 측정하고 이를 바탕으로 지원자를 초기 단계에서 스크리닝한다면 이에 꼭 들어맞지 않으나 다양한 역량과 잠재성을 가진 지원자들이 조기에 탈락될 가능성이 있다. 이는 채용업체의 조직 내 다양성을 제고하기 어렵게 만들 수 있다. 어떤 방식으로 '능력'과 '성격'이 측정되는지에 따라 채용도구는 그 개발과정에서 전형적인 것으로 규정한 것과 다른 인지적 특성이나 문화적 배경을 가진 지원자들에게 차별로 작용할 수도 있는 것이다.

국내업체들에 의해 개발, 활용되고 있는 AI 채용

도구들에 대해서도 채용업체 수준에서 유사한 분석을 시도해 볼 수 있을 것이다. 예컨대 지원자의 역량을 특정한 과제에 대한 행동반응으로부터 파악해 낼 수 있다는 전제가 타당한지, 그러한 평가 방식이 기업의 채용 목적에 부합하는지, 그리고 그 과정에서 불리한 상황에 처하게 되는 지원자들이 없는지, 이러한 배제가 기업의 다양성이나 채용의 형평성을 해치지는 않을지 등을 면밀하게 고려할 필요가 있다. 지원자의 응답 내용을 AI로 분석, 평가하는 경우에는 데이터 라벨링 과정을 보다 면밀하게 검토해야 한다. 전문면접관 라벨러의 평가 기준이 채용 목적에 부합하는지, 오랜 경험을 가진 전문가들의 평가가 오히려 그 경험에서 비롯된 편견이나 규범적 전제들을 포함하는 것은 아닐지에 대한 검토가 필요하다. 인사와 관련한 전문적 지식이 지금 급변하고 있는 기업 환경이나 조직 내 다양성, 형평성의 추구라는 새로운 비전을 담아낼 수 있을지 질문할 필요가 있다. 여러 AI 채용도구가 공통적으로 채택하고 있는 소프트스킬 분석 역시 표정이나 음성의 전형성이 소통역량에 대한 적절한 측정 자료가 될 수 있는지, 이러한 평가방식이 비전형적인 얼굴 움직임이나 목소리를 가진 지원자에게 불리하게 작동할 우려는 없는지 등에 대해서도 고려해 볼 수 있다.

이처럼 사회-기술적 매트릭스의 정성평가 항목들은 채용도구의 개발 및 설계 방식과 채용목적의 관계, 형평성의 문제를 검토하는 데 좋은 지침을 제공해 줄 수 있다. 이에 따라 AI 채용도구를 도입할 때 채용담당자가 사회-기술적 매트릭스의 문항들을 포함하는 자기기술적 정성평가 보고서를 작성할 것을 권고하는 것을 검토해 볼 수 있다. 정성평가 보고서를 작성하는 것은 AI 채용도구를 면밀하게 평가할 수 있게 할 뿐 아니라, AI 채용도구에 대한 이해 수준이 고르지 않은 채용 담당자들이 작성 과정을 통해 AI 채용 시스템에 대한 채용 담당자의 이해 수준을 높이고 도구 활용의 기대효과를 예측할 수 있게 하는 데 도움을 줄 수 있다. 이에 따라 채용도구 활용과 관련해 예상되는 문제에 대한 선제적 개입 등도 가능해질 것이다. 나아가 이렇게 작성된 정성평가 보고서의 내용을 채용업체 웹사이트에 공개함으로써 AI 채용도구에 대한 지원자의 이해를 돕는 것도 가능할 것이다. 이 자료를 통해 기업의 인사원칙 및 채용기준을 명확히 고지하는 동시에, 그것이 AI 채용도구에서 평가되는 방식을 알림으로써 지원자들이 불이익을 겪지 않게 할 수 있다.

5. AI 채용도구의 사용자 인터페이스 검토

마지막으로 알고리즘과 데이터 등에 비해 중요하게 다루어지지 않는 AI 채용도구의 사용자 인터페이스 문제를 검토하고자 한다. AI 채용도구의 형평성 제고를 위해서는 현재 인공지능과 관련한 정책에서 중요하게 다루어지고 있는 알고리즘의 신뢰성이나 공정성에 대한 검토는 물론 그것이 실제 현장에서 어떤 효과를 발생시키는지에 대한 보다 면밀한 검토가 필요하다. AI 채용도구의 사용자 인터페이스 역시 그러한 검토의 한 부분이다.

AI 채용도구의 사용자는 크게 두 그룹으로 나눌 수 있다. 먼저 AI 채용도구를 통해 평가받는 지원자 그룹이다. 지원자 그룹에게는 각각의 검사에서 측정하고자 하는 역량이 어떤 것인지에 대한 보다 구체적인 정보를 제공할 필요가 있다. 이에 더해 지원자들이 각각의 신체적, 인지적 조건을 가진 지원자들로 인해 도구 활용에 불편을 겪을 가능성이 있는지,

즉 접근성(accessibility) 문제가 없는지에 대한 고려가 필요하다. 접근성 문제는 AI 채용도구와 관련된 윤리점검표의 항목으로 포함되어 있기도 하지만 아직은 형식적인 수준에 머물러 있는 것으로 보인다. 예를 들어 게임을 기반으로 한 역량검사의 경우에는 시각장애인의 응시가 어려우며 신경다양인들에게 불리하게 작용할 수 있다는 문제가 두드러진다. 따라서 각 채용도구의 개발 및 테스트 과정에서 다양한 지원자 그룹을 포함시킬 필요가 있는 것이다. 이에 더해 채용업체는 접근성 문제로 지원자가 응시하기 어려운 경우를 대비해 대안을 제공해야 할 것이다.

또 다른 AI 채용도구의 사용자 그룹은 채용담당자들이다. 이들이 시스템에서 평가결과를 어떤 방식으로 접하게 되는지 역시 면밀히 검토해야 한다. 최근 C업체에서는 자사의 AI 역량검사를 업데이트하면서 조직부적응 가능성에 대한 판단기준을 설정하고 이와 관련한 하위 항목들을 새로 설정했다고 밝힌 바 있다. 이 업체는 이 업데이트를 하기 전에도 '부적응 가능성' 항목의 하위항목으로 '불안/우울'과 같이 정신질환을 상기시키는 아이템 등을 포함시킨 바 있었다. 2024년의 업데이트에서는 '꼰대', '번아웃', '뒷담화', '분노조절장애' 등 일상용어들을 사용해 "입사 후 모습을 직관적으로 예측할 수 있도록" 하는 하위항목들을 설정했다는 것이다.

이러한 일상용어들은 직관적이기 때문에 오히려 채용 과정에서 젠더나 연령에 대한 고정관념과 쉽게 결합될 우려가 있다. 예를 들어 번아웃과 분노조절장애는 통념적으로 각기 여성과 남성에 결부되는 경향이 있다. 이 때문에 여성 지원자의 검사 결과에서 '번아웃' 점수가 높게 나왔다면 면접관은 이를 더 신빙성 있는 결과로 받아들일 수도 있고, 이와 반대로 여성 지원자가 '분노조절장애' 점수를 받은 경우에는 그것이 예외적이라고 여겨지기 때문에 더 심각한 문제로 해석할 가능성도 있다. 나이가 상대적으로 많은 지원자에 대해 AI가 '꼰대'라고 평가했다면 이 결과지는 지원자의 연령과 관련한 편견을 강화할 수도 있다. 이러한 검사결과는 면접관이나 인사 관계자가 참고자료로 활용하는 과정에서 기존의 성별, 연령과 관련한 통념이 미칠 영향을 고려해야 한다. 이처럼 AI 채용도구의 형평성을 확보하기 위해서는 데이터의 수립, 처리 과정 등과 알고리즘의 작동 방식뿐 아니라 검사결과에서 사용되는 언어, 평가 항목, 시각화 방식까지 면밀히 검토할 필요가 있다.

6. 나가며

AI 채용도구와 관련된 지금까지의 논의에서는 주로 데이터의 편향, 알고리즘의 공정성 등 AI 개발과정의 문제가 주요하게 다루어져 왔다. 본고에서는 이러한 논의에서 간과됐던 문제, 즉 현장에서 실제로 AI 채용도구가 도입, 적용될 때 발생할 수 있는 문제들에 주목할 것을 제안했다. 채용에서의 형평성은 단순한 기계적 공정성에 관한 것이 아니다. 형평성 제고를 위해서는 지원자들 사이의 차이에 대한 충분한 고려, 나아가 사회구조적인 이유로 취약한 위치에 있는 지원자들이 자신의 역량을 발휘할 수 있는 조건을 만들 필요가 있다. 따라서 AI 채용도구의 형평성을 제고하기 위한 전략은 개발 과정에 대한 고려뿐 아니라 그것이 활용되는 과정의 구체성과 그 효과 등에 대한 검토를 필요로 한다.

• 참고문헌 •

과학기술정보통신부·정보통신정책연구원(2023). 인공지능 윤리기준 실천을 위한 자율점검표(안). 과학기술정보통신부·정보통신정책연구원.

과학기술정보통신부·한국정보통신기술협회(2024). 2024 신뢰할 수 있는 인공지능 개발 안내서 – 채용 분야. 과학기술정보통신부·한국정보통신기술협회.

인사혁신처(2018). 인사혁신 사례집. 인사혁신처.

Equal Employment Opportunity Commission. DEMOGRAPHIC INFORMATION ON APPLICANTS. https://www.eeoc.gov/sites/default/files/2021-01/Applicant_data_OMB_July_2020.pdf(최종 접속일: 2024.11.24.)

Sloane, M., Moss, E., & Chowdhury, R. (2022). A Silicon Valley love triangle: Hiring algorithms, pseudo-science, and the quest for auditability. *Patterns,* 3(2), pp. 1-9.