"Trustworthy AI (AI 신뢰성)" 특집호 발간에 즈음하여

최근 인공지능(AI) 기술은 행정, 금융, 의료, 국방 등 국가와 사회 전반에 걸쳐 빠르게 확산되고 있으며, 업무 혁신과 효율화에 큰 기여를 하고 있습니다. 그러나 동시에 AI가 만들어 내는 결과물에 대한 신뢰성 문제와 보안 위협이 새로운 도전 과제로 떠오르고 있습니다. 이에 따라 국내외적으로 신뢰할 수 있는 AI(Trustworthy AI)에 대한 연구와 정책 논의가 활발히 전개되고 있으며, AI 시스템을 안전하고 책임 있게 운영하기 위한 기반 기술과 제도적 장치의 필요성이 커지고 있습니다.

기존의 AI 기술 활용은 주로 성능 향상과 편의성 증대에 초점이 맞춰져 있었으나, 이제는 공정성, 설명 가능성, 투명성, 안전성과 같은 가치가 동시에 고려되지 않으면 사회적 수용성을 확보하기 어렵다는 인식이 확산되고 있습니다. 특히 AI가 보안 영역과 접목될 경우, 공격자의 위협 수단으로 악용될 가능성뿐 아니라 AI 시스템 자체가 침해당할 위험까지 존재하기 때문에, AI 보안은 단순한 기술 문제가 아니라 사회적 신뢰와 직결된 중요한 과제가 되고 있습니다.

이번 특집호에서는 금융 분야에서의 AI 활용 및 보안 기술 동향, AI 기반 사이버공격 대응 전략, 딥페이크 생성 및 탐지기술 동향, 딥러닝의 취약점 탐지 및 한계, Agent AI 시대의 공격·방어 기술 동향, AGI 보안이슈 및 전망 등 다양한 주제를 다루었습니다.

이 특집호가 AI 활용 확산 속에서 신뢰성과 보안성을 동시에 확보해야 하는 국가·산업·학계의 고민에 의미 있는 참고 자료가 되기를 기대합니다. 끝으로, 바쁘신 와중에도 귀중한 원고를 집필해 주신 필진 여러분과 학회지 발간을 위해 힘써 주신 편집위원회 및 학회사무국 관계자 여러분께 깊은 감사를 드립니다.

2025년 10월

가천대학교 스마트보안학과 이 태 진