

□ 기획연재 □

컴퓨터 과학 산책(11)

아시모프와 로봇

숙명여자대학교 이만재*

컴퓨터가 생활의 일부로 되어 있는 사람들에게 로봇이라는 단어는 비교적 친숙한 단어이다. 로봇은 “스타워즈”의 R2D2나 C3PO와 같이 인간을 도와주는 로봇도 있으나 “터미네이터”에서 등장하는 로봇처럼 인간의 가장 무서운 적이기도 하다. 이러한 로봇의 기계적 움직임은 기계공학자의 몫이 될 수 있으나 로봇의 두뇌는 정보과학을 하는 우리들의 영역이기에 과연 터미네이터에 등장하는 로봇이 가능한가를 살펴 본다.

로봇이 처음 등장한 것은 1812년 매리 셸리의 “프랑켄스타인”을 손꼽는다. 이 소설은 미친 과학자와 로봇이라는 틀에 박힌 형식을 처음 보여주기도 하였다. 그러나 셸리의 소설에는 로봇이라는 단어가 사용되지는 않는다. 로봇이라는 단어는 체코어에서 강제로 일하는이란 뜻을 가진 robota에서 출발하였으며 카렐 카פק의 희곡 Rossum's Universal Robot에 처음 등장하였다. 이 희곡에서는 로섬이라는 영국인이 로봇을 대량으로 생산하여 인간의 노동을 대신하도록 하였으나 결국 로봇은 인간에게 반란을 일으켜 인간을 멸망시킨다는 이야기로 되어 있다.

글을 이어나가기 전에 이 글의 대상이 되는 로봇을 아래와 같이 정의하고자 한다. 첫째 로봇은 프로그램이 가능하여야 한다. 즉 인간의 두뇌를 그대로 사용하고 신체의 일부를 기계로 바꾼 “로보캅”이나 “6백만불의 사나이”는 로봇이 아니다. 다시 말하면 로봇의 두뇌는 컴퓨터로 이루어져 있어야 한다.

두번째 로봇은 기계적으로 환경에 적응할 수 있어야 한다. 단순히 가상적으로 존재하는 프로그램은 로봇이라 하지 않는다. 인공지능 연구에 등장하는 프로그램은 개발자와의 대화에서는 지능을 발휘하나 물리적인 힘을 사용하지는 못한다. 비록 본체는 움직일 수 없다고 하더라도 주변의 물체를 제어하여 기계적인 일을 할 수 있을 경우에 로봇이라 한다.

세번째로 로봇은 비교적 광범위한 일을 할 수 있어야 한다. 한 가지 일만 할 수 있는 지능 있는 기계는 하나의 기계일 뿐이다. 자동차 조립현장에서 용접을 한다든지 하는 위험한 일을 사람을 대신하는 경우에는 다양성 면에서 로봇이라 보기 어렵다.

이러한 로봇에 대한 최초의 학문적인 접근은 로보틱스(robotics)라는 단어로부터 출발한다. 로보틱스는 공상과학소설의 대가 중 한 사람인 아이작 아시모프에 의해 1942년 처음 사용되었다. 아시모프의 소설에는 로봇이 자주 등장한다. 그의 소설에 등장하는 로봇은 인간을 도와 주는 역할을 할 뿐이지 인간에 대해 반란을 일으키는 등 독자를 흥분시키는 행동을 하지 않는다. 아시모프는 로봇이 지켜야 할 3대 법칙을 정하였고 그의 소설에 등장하는 모든 로봇은 모두 이 법칙을 따르도록 하였다. 이 법칙을 옮기면 다음과 같다.

제1법칙 : 로봇은 사람을 해치지 않아야 하며 사람이 위험에 처할 때 대처하지 않아 사람에게 해를 입혀서는 안 된다.

제2법칙 : 로봇은 제1법칙에 어긋나지 않는 한 사람의 명령을 따라야 한다.

제3법칙 : 로봇은 제1법칙과 제2법칙에 어긋

‘나지 않는 한 자신을 보호하여야 한다.

이 법칙을 어떻게 로봇의 두뇌에 프로그램하느냐는 방법론은 뒤로 미루더라도 이 3가지 법칙이 가르치는 바는 크다. 제1법칙이 지켜진다면 인간을 해치는 “터미네이터”나 “2001: 스페이스 오디세이”에 등장하는 HAL이라는 로봇은 불가능하다. 제2법칙은 인간이 로봇을 만든 가장 큰 이유이다. 즉 로봇을 시켜 일을 하라는 목적을 나타낸다. 제3법칙은 대부분의 사람이 간과할 수 있는 법칙이다. 인간이 만든 로봇을 역사에 비추어 보면 중세 이전의 노예라고 볼 수 있다. 노예는 자기의 신세를 한탄하고 때에 따라서는 자살하고 싶은 충동을 느낄 수 있다. 아시모프는 로봇이 같은 심정을 느끼고 행동한다면 이를 막을 장치를 필요로 하였으며 이를 제3법칙에서 생존의 필요성을 강조함으로써 해결하였다.

영화 터미네이터에 등장하는 로봇은 흥행 위주로 만든 작품이라 제외하면 아시모프와 비견되는 공상과학 소설의 대가이며 과학자인 아서 클락의 “스페이스 오디세이 2001”에서 우주선을 통괄하는 컴퓨터인 HAL은 왜 승무원을 해쳐야 했나 하는 것은 의문으로 제기된다. 단순한 컴퓨터의 고장인가? 아니면 더 큰 수수께끼가 숨어 있는가? 아서 클락은 속편인 “2010: 스페이스 오디세이”에서 이 문제를 해결한다. HAL은 우주선의 출발에 앞서 지구의 사령부에서 목적에 있을지도 모르는 외계문명을 탐사하라는 명령을 받는다. 그리고 이 명령은 다른 명령에 우선하는 것이며 도착하기 전까지는 우주선의 원래 목적이 지구에 알려지는 것을 막기 위해 승무원에게도 알려지 말 것을 또한 명령받는다. HAL은 자기가 부여받은 가장 중요한 임무를 우선적으로 수행하기 위해 인간을 해치기 시작하나 결국 주인공과의 싸움에서 지고 기능을 정지당한다. 결국 인간이 HAL에게 제1법칙을 프로그램으로 구현하지 않은 오류를 범한 것으로 결론짓는다.

아시모프의 소설에 등장하는 로봇은 이 3가지 법칙을 지키기 위해 여러가지 어려움을 겪는다. 우선 인간에게 어떤 것이 위험한 지를 아는 것이 어렵다. 로봇의 힘은 인간보다 훨씬 세기 때문에 이를 기준으로 생각하면 인간은

너무 연약한 존재이며 이러한 특성을 알지 못하면 인간을 위협에서 구할 수 없다. 인간을 어떻게 구별하는가 하는 문제는 한층 어려운 숙제이다. 인간의 피부색, 언어, 음성은 매우 다양하며 잘못하면 마네킹을 인간으로 착각할 수도 있기 때문이다. 성인 백인을 인간의 판단으로 기준 삼는다면 아프리카의 피그미족 인간은 너무나 인간의 기준에서 떨어져 있다. 이러한 문제를 풀어 나가는 과정에서 인간을 호모 사피엔스에 속하는 종으로 정의한다.

그러나 아시모프의 세 가지 법칙은 더 큰 도전을 받는다. 사람을 구하기 위해서는 어느 정도 인체에 해를 가해야 하는 상황이 발생하게 된다. 나아가서는 두 사람이 생명의 위협을 받고 있고 그 중 한 사람만을 구할 가능성 밖에 없을 경우 누구를 택해야 하는가 하는 문제이다. 예를 들면 다음과 같은 상황이 발생한다. 한 사람이 로봇에게 “저 사람을 죽이지 않으며 나는 자살하겠다”라는 어려운 명령을 지시한다. 이러한 문제를 푸는 과정에서 한 로봇은 로봇의 3대법칙에 문제점이 있음을 발견하고 “로봇이 쓴 로보틱스”라는 저서를 출판하며 제1법칙은 다음과 같이 수정된다.

제1법칙(수정): 로봇은 인간의 최종적인 안전을 위한 다른 대안이 없는 경우를 제외하고는 인간에게 해를 끼치지 않아야 한다.

이러한 수정 법칙에 따라 인간을 구하기 위해 어느 정도의 위험부담을 감수할 수 있게 된다. 그러나 이 수정된 법칙도 위의 두번째 문제를 풀지 못하고 대드록 상태에 빠진다.

결국 단순한 법칙의 수정만으로는 해결할 수 없음이 발견되고 아시모프는 1985년 3가지 법칙에 우선하는 제0법칙을 추가한다.

제0법칙: 로봇은 인류에게 해를 끼치지 않아야 하며 인류가 위협에 처할 때 대처하지 않아 인류에게 해를 입히지 않아야 한다.

이 경우에도 물론 제1법칙 이하는 상위 법칙이 우선함을 전제로 한다. 제0법칙은 인간이 완전하지 못함을 법칙으로 인정한 셈이며 불완전한 인간들의 집합이지만 인간이 총체적으로 추구하는 목적을 기준으로 인류라는 개념을 사용하였다. 그러나 이 새로운 법칙은 더 큰 모순을 발견하게 된다. 인류라는 특성을 가진 인

간은 과연 존재인가 하는 문제이다. 다른 사람의 행복을 위해 일하는 애쓰는 사람을 모범적인 인간으로 존중한다. 그런데 과연 로봇만큼 인간을 위해 애쓰는 것이 있을까, 즉 로봇이 가장 인간적인 것은 아닐까 하는 의구심이다. 이러한 의구심을 품은 로봇은 가장 인간다운 것은 로봇임을 확신한다.

아시모프의 소설은 그러나 인간이 되고 싶어 하는 로봇을 주제로 하고 있지만 인간을 대신 하려는 로봇은 등장시키고 있지 않다. 인간이 되고 싶은, 그리고 인간보다 더 인간적인 삶을 사는 한 특이한 로봇은 인간으로부터 따돌림을

받는다. 인간의 수명보다 오래 사는 로봇을 인류의 한 구성원으로 대할 수 없기 때문이었다. 결국 너무나 인간이 되고 싶은 로봇은 자기 두뇌의 논리 회로에 망각기능을 포함시키도록 하여 스스로 수명을 단축시키고 인류의 일원으로 인정받는다. 그 소설 이후 아시모프는 더 이상 로봇에 대한 이야기를 전개시키지는 못하고 1992년 4월 6일 세상을 떠났다. 그러나 그가 남긴 로봇의 이야기는 인공지능을 개발하여 로봇을 만들라는 학자에게 이 일이 얼마나 어려운 일인지를 소설 속의 가상현실로 말하고 있다.

● 제23회 정기총회 및 추계학술발표회 ●

- 행사일정 : 1996년 10월 25(금)~26일(토)
- 행사장소 : 한국외국어대학교(용인)
- 발표논문 접수마감 : 1996년 8월 24일(토)
- 문의 및 접수처 : 한국정보과학회 사무국

Tel. 02-588-9246, Fax. 02-521-1352

서울시 서초구 방배3동 984-1(머리재빌딩) ☎137-063

● APSEC '96 ●

- 일 자 : 1996년 12월 4~7일
- 장 소 : 교육문화회관
- 주 최 : 소프트웨어공학연구회
- 문 의 처 : 포항공과대학교 강교철 교수

T. 0562-279-2258

F. 0562-279-2299

E-mail: kck@wision.postech.ac.kr