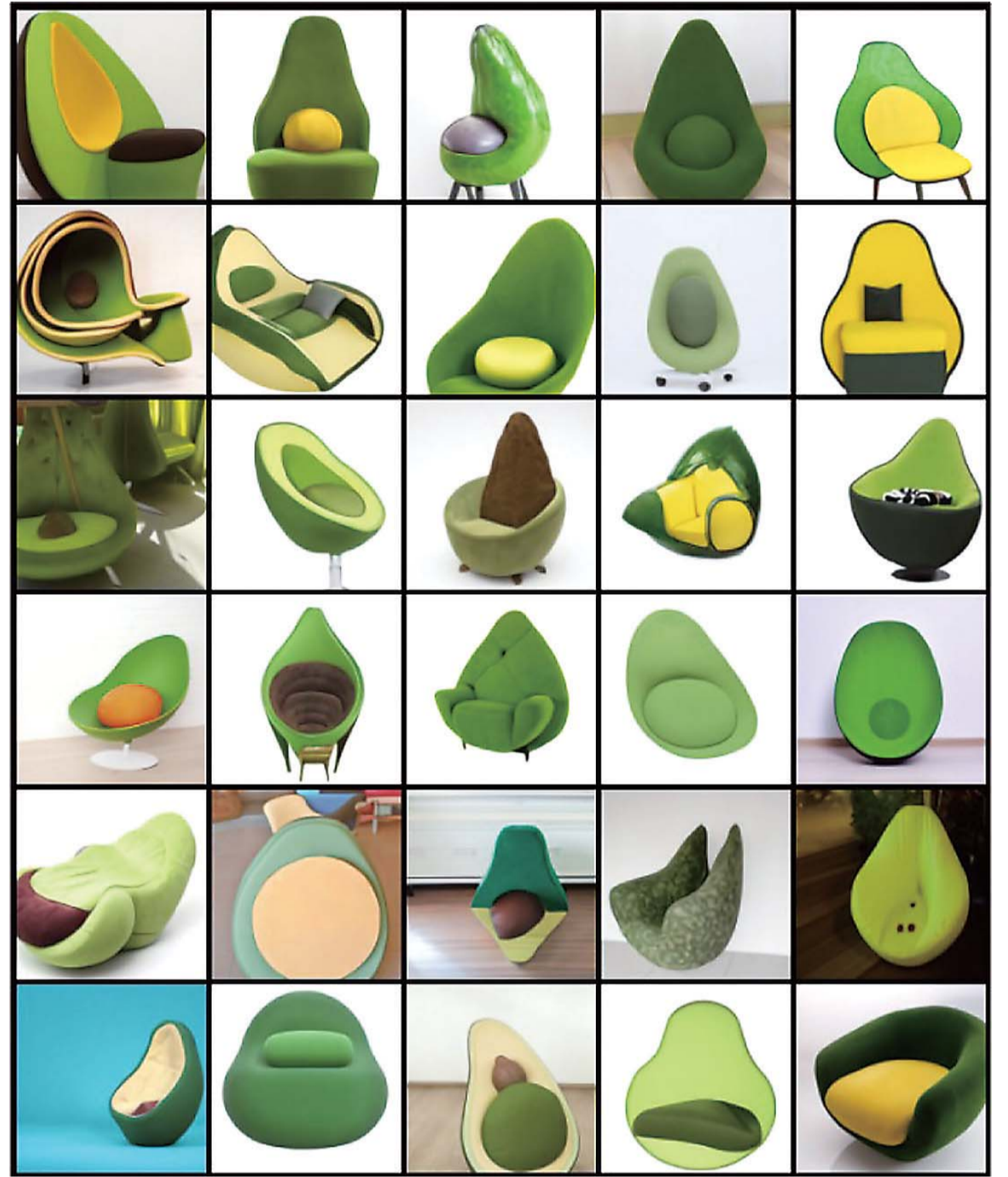
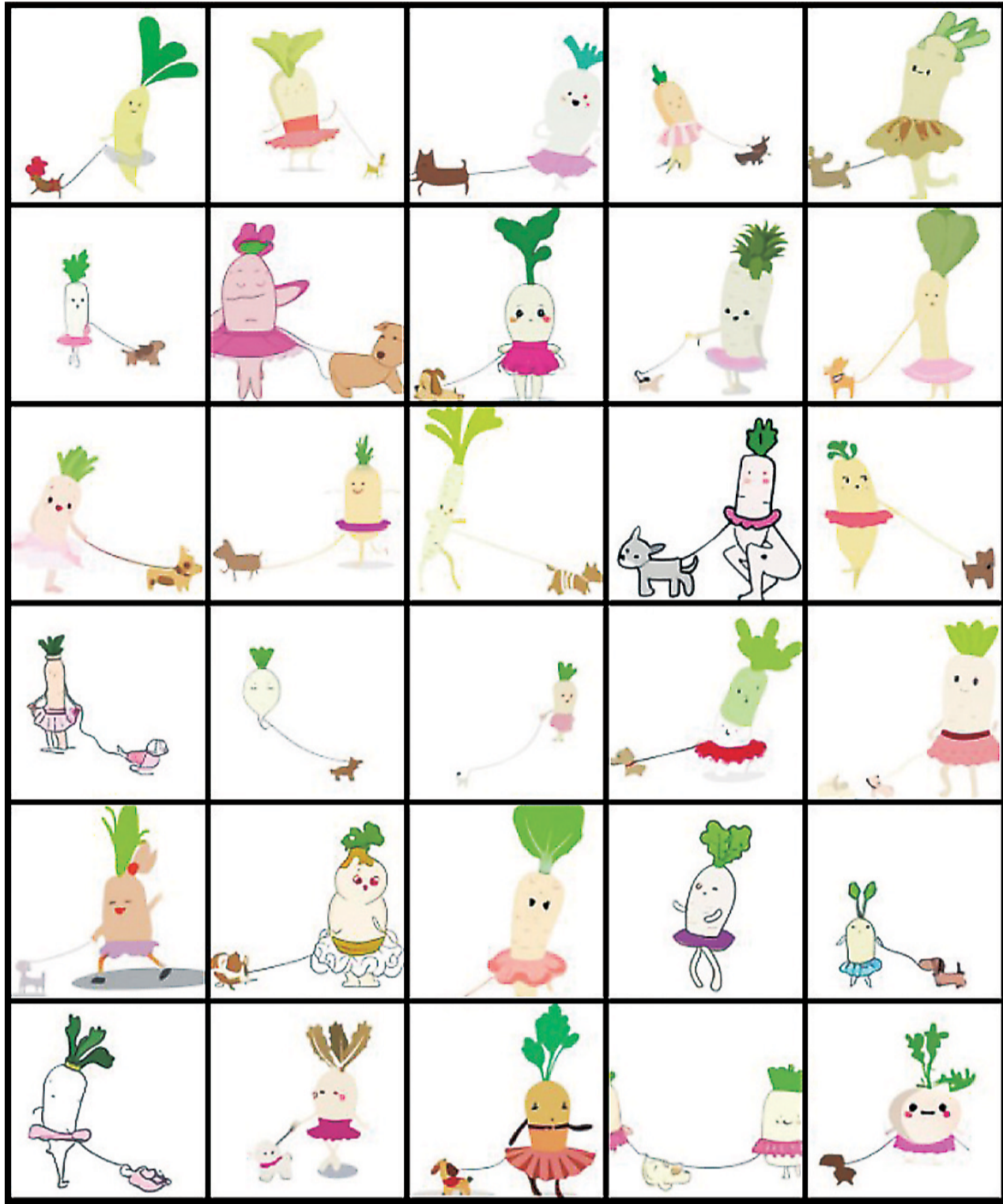


어떤 그림도 척척 ... 인공지능, 달리 솜씨 못잡네

〈초현실주의 화가〉



〈사진 = 오픈AI 캡처〉

‘발레 치마를 입은 아기 무가 강아지와 산책하는 모습’. 어린이이라도 아마 이 글이 묘사하는 장면을 그림으로 그릴 수 있을 것이다. 잘 그릴 수 있는지는 다른 문제지만 말이다. 하지만 이 짧은 글에 묘사된 모습을 그릴 수 있다는 것은 인간이 여러 신기한(?) 능력을 복합적으로 가지고 있음을 알려준다. 무라는 식물을 전혀 다른 존재인 사람의 모습에 빗대 머릿속에 그림 수 있어야 한다. 무가 옷을 입고 강아지와 함께 다닌다는, 살면서 한 번도 본적 없는 모습을 생각만으로 재현해야 한다.

전혀 상관없는 사물이나 개념들을 한데 묶어 새로운 것을 제시하는 능력, 그것을 창의성이라고 부른다. 완전히 새로운 것을 무(無)에서 만들어내야 창의적이라는 세간의 생각은 미신에 가깝다. 그래서 인공지능이 실제 지능이 있는지를 판정하기 위해 이렇게 여러 개념을 뒤섞어 새로운 것을 보여주는 능력을 시험해 보자는 아이디어를 낸 사람이 있다. 마크 리틀 조지아공대 교수다. 그는 2014년 ‘라브레이스 2.0’이라는 테스트를 제안했다. 이를테면 ‘뽕글을 든 남자’를 그릴 수 있다면 그 컴퓨터는 지능이 있다고 볼 수 있다는 것이다.

앨런 튜링은 1950년 사람이 자기가 (채팅으로) 대화하는 상대가 사람인지 컴퓨터인지 분간할 수 없다면 그 컴퓨터는 지능이 있다고 볼 수 있다며 ‘튜링 테스트’를 제안했다. 리틀 교수는 이를 발전시켜, 사람이 그린 것과 분간할 수 없는 그림을 그리는 컴퓨터는 지능이 있다고 볼 수 있다고 주장한 셈이다.

설명만 듣고 사람이 그린 것과 분간할 수 없는 그림을 그리는 인공지능이 가능할까? 지난해 사람이 쓴 것과 구분할 수 없는 글을 써내는 인공지능 모델 GPT-3를 발표해 세상을 놀라게 했던 미국의 인공지능 연구기관 오픈AI가 이번에는 그림으로 우리를 또 한 번 놀라게 했다.

GPT-3는 엄청난 양의 텍스트를 학습해 사람처럼 그럴듯한 글을 써낼 수 있도록 만든 인공지능 모델이다. 인터넷에서 수집한 4990억개의 텍스트 데이터를 교재로 학습했다. 어떤 단어를 접했을 때 그다음에 나올 가장 적절한 단어를 예측하는 방식이다. GPT-3가 학습한 4990억개의 텍스트에는 아마 세계의 삼라만상이 거의 다 들어있을 것이다. 그래서 GPT-3는 시인보다 더 시인 같고, 기자보다 더 기자 같은 글을 토해냈다. 철학자보다 더 철학적인 대화를 나누었다. 실제로는 글을 이해하지 못하는데도 말이다.

오픈AI는 이 접근법을 그림으로 확대했다. 텍스트 대신 디지털 이미지를 이루는 화소, 즉 픽셀로 인공지능을 훈련하는 ‘이미지 GPT-3’ 프로젝트였다. GPT-3가 단어를 보고 다음 단어를 예측하듯, 이미지 GPT-3는 반쯤 그려진 그림을 완성할 수 있었다.

인공지능 모델 GPT-3 만든 오픈AI 텍스트·이미지 익힌 ‘DALL-E’ 내바

‘아보카도 안락의자’ 등 금세 형상화 창의성 더 키우는 도구로 활용할 만



그리고 여기서 한발 더 나아가 최근 ‘DALL-E’와 ‘CLIP’이라는 새로운 이미지 관련 인공지능 모델을 선보였다. 이미지와 텍스트를 모두 사용해 훈련한 인공지능 모델이다. 인터넷에서 수집한 이미지와 그 이미지에 딸린 설명을 학습했다. 통상 인공지능의 훈련에 쓰이는 데이터 세트(Dataset)의 이미지는 ‘고양이’ ‘자동차’ 등의 꼬리표가 달려 있고, 인공지능 모델은 이런 데이터로 학습한다. 하지만 이들 오픈AI 인공지능 모델은 사진에 대한 설명을 보고 학습했다.

DALL-E는 초현실주의 화가 살바도르 달리와 인공지능 로봇이 나오는 영화 ‘월-E’를 패러디해 붙인 이름이다. 시계나 탁자 같은 일상의 사물을 기괴하고 새로운 모습으로 묘사한 달리의 그림처럼, DALL-E도 익숙한 개념을 새로운 방식으로 섞어서 보여주는 능력을 보여준다. 결과물도 훌륭한 편이다. ‘하프루 만든 달팽이’나 ‘아보카도 안락의자’ 같은 문장을 주면 그럴 듯한 그림을 그

려낸다.

GPT-3가 ‘텍스트 생성기’로서 놀랄 만한 성과를 보여줬다면, DALL-E는 ‘이미지 생성기’로서 뛰어난 성능을 보여준다. 특히 텍스트를 갖고 이미지를 만들어낸다는 점이 특징이다. ‘오각형 모양의 녹색 탁상시계’나 ‘테이블 위에 놓여진 술잔들’, ‘호두의 단면도’ 같은 설명이 묘사하는 장면을 정확히 그려낸다.

‘석양이 지는 들만에 앉아 있는 카피라(토끼를 닮은 대형 설치류로 남미에 주로 서식한다)의 그림’이라는 문장을 주면, 카피라의 위치나 바라보는 방향, 태양의 위치 등을 따져 그림자를 그린다. 설명이 명시적으로 언급되지 않은 사항도 추정해 그리는 것이다.

DALL-E가 설명을 읽고 이미지를 그린다면, CLIP은 연관이 있는 텍스트와 이미지를 연결한다. 묘사된 내용과 가장 일치하는 이미지를 찾아내는 이미지 인식 기술인 셈이다. 3만2768개의 무작위 사진 설명 중 주어진 이미지

에 가장 적합한 것은 무엇인지 예측하도록 훈련된다. 정교하게 분류되지 않은 데이터로도 효과적인 학습이 가능하게끔 했다. 이를테면, 얼굴 인식 인공지능을 만들려면 고품질의 얼굴 데이터로 추가적 훈련을 시켜야 했지만, 이 모델은 추가 훈련 없이 다양한 과제를 수행할 수 있다는 말이다.

텍스트만 갖고 훈련하던 GPT-3와 달리, DALL-E와 CLIP은 텍스트와 이미지를 함께 사용해 학습한다. 우리가 눈에 보이는 것에 이름을 붙여 세상을 이해하듯 인공지능도 시각과 텍스트로 효과적으로 세상을 배울 수 있는 길이 열린 것이다. 인공지능이 인간의 세계를 더 잘 파악할 수 있게 되어 인공지능의 활용 여지가 더 커질 것이란 기대를 일으킨다.

또 DALL-E는 인공지능이 인간의 창의성을 얼마나 대체할 수 있는가에 대한 질문을 던진다. 그간 간단한 작업임에도 디자이너나 일러스트레이터가 없어 하지 못했던 일들을 이제는 할 수 있으리라는 기대가 들기도 한다. 인공지능에 글로 설명만 써 주면 간단한 결과물은 얻을 수 있는 것이다.

전혀 상관없어 보이는 엉뚱한 개념을 엮어 참신한 결과물을 내놓는 것이 사람만의 영역이라고 주장하기도 어려워질 듯하다. 인공지능은 과거의 데이터를 학습해 결과를 내놓기 때문에 새로운 것을 만드는 창의적 작업에는 적합하지 않다는 말을 흔히 한다. 하지만 충분히 많은 양의 데이터를 인공지능에 ‘먹인’ 결과, 우리는 하프루 만든 달팽이나 아보카도 안락의자 같은 비현실적이고, 아마도 인공지능이 전에 들어본 적 없었을 내용을 가지고도 훌륭한 결과물을 만들어내는 현실을 보고 있다.

물론 이들 인공지능 모델도 GPT-3와 같은 한계는 그대로 갖고 있다. 많은 결과물이 훌륭하지만, 어색하거나 잘못된 결과물도 여전히 나온다. 인공지능이 실제로 말이나 글을 이해하고 결과를 내는 것이 아니라는 점도 같다. 그리고, 인공지능의 발달이라는 현실을 만난 우리는 겁먹기보다는 이 도구를 창의성을 더 키우는 쪽으로 활용해야 한다는 사실도 변함없다.

/한세희 IT칼럼니스트

한세희 필자는 전자신문 기자와 동아시아인스 데일리뉴스팀장을 지냈다. 기술과 사람이 서로 영향을 미치며 변화하는 모습을 항상 흥미진진하게 지켜보고 있다. ‘어린이를 위한 디지털과학 용어 사전’을 지었고, ‘네트워크전쟁’을 옮겼다.

〈광주일보와 중앙SUNDAY 제휴 기사입니다〉