

무료 공개 이북 · 챕터 2



AI 브라우저

블로그 자동화, 강의 말고 원리로 배웁니다

정원석 지음 · Connect AI LAB

이 챕터에서

1	강의를 사기 전에 손이 아니라 원리를 남깁니다
2	AI 브라우저란 확장 프로그램이 아닙니다
3	만든 사람들 한국인 세 명이 만들었습니다
4	원리 하나 화면이 아니라 설계도를 읽습니다
5	원리 둘 단추를 누르지 않고 코드를 씁니다
6	원리 셋 크로미움을 가져다 개조했습니다
7	원리 넷 AI 를 만들지 않았습니다
8	루프 생각하고, 행동하고, 본다
9	논문 하나 — WebGPT 브라우저를 켜 첫 언어모델
10	논문 둘 — ReAct 루프에 이름을 붙이다
11	논문 셋 — Mind2Web 실제 웹으로 만든 시험지
12	99%, 누가 잤나 숫자를 읽는 법
13	로컬 AI 완전 무료, 완전 소유
14	안전 임시저장까지만
15	낱말 풀이 이 책에 나온 말들

각 절 끝에 강의 영상이 붙어 있습니다. QR 을 찍으면 휴대폰에서도 열립니다.

1

강의를 하기 전에

손이 아니라 원리를 남깁니다



몇십만 원짜리가 팔리고 있습니다

「수익화 블로그 자동화」라는 이름으로 강의가 팔립니다. 몇십만 원씩 합니다.

그런데 그 안에서 가르치는 것의 상당 부분이 **지금 무료로 되는 일**입니다. 이 책은 그걸 직접 해 보시라고 쓴 것입니다.

도구는 바뀌고 원리는 남습니다

강의를 파는 쪽은 대개 손을 보여 줍니다. 여기를 누르고 저기를 누르라고요. 그 도구가 6개월 뒤에 바뀌면 배운 것이 통째로 날아갑니다.

이 책은 원리를 남깁니다.

1

무엇을 읽는가

화면이 아니라 웹페이지의 설계도

2

어떻게 조종하는가

단추가 아니라 코드

3

무엇으로 만들어졌는가

크로미움을 개조한 것

4

두뇌는 어디서 오는가

만들지 않고 빌려 온다

이 네 가지를 알면 다른 도구가 나와도 첫 화면에서 바로 파악하실 수 있습니다.

이 책에 적힌 숫자와 사실은 회사 기술 블로그·홈페이지와 논문 원문에서 직접 확인한 것만 실었습니다.

2

AI 브라우저란

확장 프로그램이 아닙니다



브라우저 자체입니다

지금까지 나온 AI 자동화 도구는 대개 **확장 프로그램**이었습니다. 크롬 위에 얹혀서 일부만 건드렸습니다.

오늘 보는 것은 다릅니다. **브라우저 자체가 AI 입니다.** 이름은 **Aside** 입니다.

무엇이 다른가

항목	내용
로그인	내가 이미 로그인해 둔 사이트에서 그대로 일합니다. 계정 연동도, API 신청도 없습니다
지속	몇 분이든 몇 시간이든 이어서 합니다
가시성	무엇을 눌렀는지 화면에 그대로 보입니다
범위	메일·대시보드·사내 도구·문서·스프레드시트

「로그인을 상속받는다」가 핵심입니다. 그래서 **API 를 열어 주지 않는 사이트에서도** 일이 됩니다. 네이버 블로그 같은 곳 말입니다.

강의에서 파는 「블로그 자동화」의 실체가 대개 이 구조입니다.

3

만든 사람들

한국인 세 명이 만들었습니다

세 명, 그리고 와이콤비네이터

거대 자본이 만든 것이 아닙니다. **한국인 창업자 세 명**이 만들었습니다.

본사는 샌프란시스코, 사무실은 서울에 있습니다. 와이콤비네이터를 거쳤습니다.

이 사실이 왜 중요한가

「세계 최고 수준의 도구는 거대 기업만 만든다」는 생각이 우리를 멈춰 세웁니다.

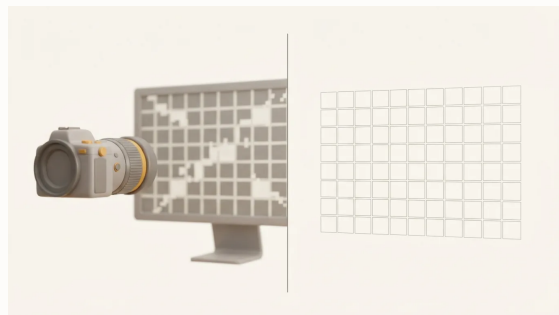
세 명이 만든 것이 지금 이 분야에서 가장 높은 점수를 주장하고 있습니다. **규모가 아니라 구조를 잘 잡은 쪽이 이깁니다.**

그 구조가 무엇인지가 4장부터 7장까지의 내용입니다.

4

원리 하나

화면이 아니라 설계를 읽습니다



흔한 방식 — 화면을 사진으로 봅니다

컴퓨터 유즈라는 말을 들어 보셨을 겁니다. AI 가 화면을 **사진으로 찍어서** 보고, 「여기쯤 단추가 있네」 하고 **좌표를 찍어 누르는** 방식입니다.

사람이 하는 것과 가장 비슷해 보입니다. 그런데 문제가 셋입니다.

1

비쌉니다

사진 한 장이 글자 수천 자에 해당합니다. 단계마다 찍으면 비용이 쌓입니다

2

느립니다

찍고, 보내고, 분석하고, 좌표를 정합니다

3

틀립니다

화면이 조금만 움직여도 좌표가 어긋납니다

이 브라우저 — 설계를 읽습니다

웹페이지에는 눈에 보이는 화면 말고 **설계도**가 따로 있습니다. 「여기는 제목, 여기는 입력칸, 여기는 저장 단추」 하고 **글자로 적혀 있는** 구조입니다. 원래는 시각장애인용 화면 낭독기를 위해 만들어진 것입니다.

접근성 트리 (Accessibility Tree)

AI 는 이것 읽습니다. 사진이 아니라 **글자**를 읽으니 싸고, 빠르고, 정확합니다.

이 회사는 설계를 자체적으로 다듬어 남들 것보다 70% 작게 줄였다고 밝혔습니다. 읽을 것이 적을수록 싸고 빠릅니다.

그러면 사진은 언제 쓰나

방식	읽는 것	언제
주력	접근성 트리 (글자)	대부분의 웹페이지
보조	스크린샷 (사진)	설계도로 제어가 안 되는 화면

사진 방식을 버린 것이 아니라 뒤로 미뤄 둔 것입니다. 이 순서가 비용과 정확도를 가릅니다.

5

원리 둘

단추를 누르지 않고 코드를 씁니다

사실은 코딩 에이전트입니다

보통 이런 도구는 AI 에게 도구 목록을 쥐여 줍니다. 「클릭하기」, 「글자 입력하기」, 「스크롤하기」 같은 것들입니다. AI 는 그중에서 골라 씁니다.

이 브라우저는 다릅니다. **조종 코드를 그 자리에서 직접 씁니다.**

이 에이전트는 사실 코딩 에이전트입니다. — 회사 기술 블로그

왜 그게 더 나은가

도구 목록에서 고르는 방식은 **한 번에 하나씩**입니다. 열 단계짜리 일이면 열 번을 왕복합니다.

코드를 쓰면 **열 단계를 한 번에** 적어 실행할 수 있습니다. 조건도 넣고, 반복도 넣습니다.

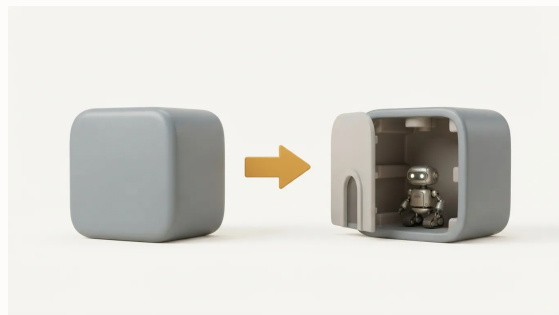
Playwright 문법의 자바스크립트를 브라우저 안에서 실행

Playwright 는 원래 개발자들이 웹사이트를 자동으로 시험할 때 쓰던 도구입니다. 이미 검증된 문법을 AI 에게 그대로 쓰게 한 것입니다.

6

원리 셋

크로미움을 가져다 개조했습니다



크롬의 뼈대는 크롬이 아닙니다

크롬의 속은 크로미움이라는 오픈소스입니다. 구글이 공개해 두었고 누구나 가져다 쓸 수 있습니다.

브라우저	뼈대
크롬	크로미움
엣지	크로미움
네이버 웨일	크로미움
이 브라우저	크로미움

「빼낸 것」이 아니라 「개조한 것」

흔히 「크롬에서 그 부분만 빼서 만들었다」고 이해하시는데, 정확히는 반대입니다.

통째로 복제한 다음(포크) 안을 고쳤습니다. AI 직원이 살 수 있는 집으로 개조한 것입니다.

1

에이전트용 홈

AI가 브라우저 안쪽을 직접 만질 수 있게 길을 냈습니다

2

봇 감지 대응

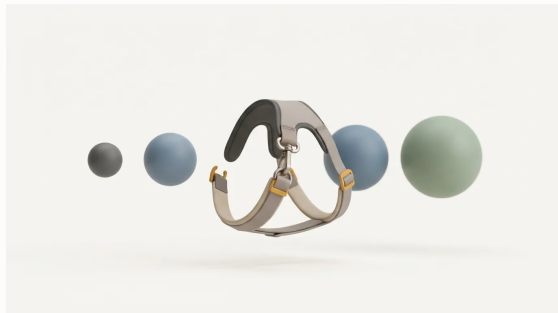
사이트가 「이건 사람이 아니다」라고 막지 않도록 고쳤습니다

※ 포크(fork)는 오픈소스를 통째로 복제해 자기 방향으로 고쳐 가는 것을 말합니다. 원본은 그대로 남습니다.

7

원리 넷

AI 를 만들지 않았습니다



두뇌는 빌려 오고, 마구를 만들었습니다

이 회사는 인공지능 모델을 만들지 않았습니다.

두뇌는 GPT 나 클로드처럼 이미 있는 것을 연결해서 씁니다. 자기들이 만든 것은 그 두뇌가 브라우저를 다루게 해 주는 장치입니다. 이걸 하네스(harness), 우리말로 **마구**라고 합니다.

말에 안장과 고삐를 채우는 그 마구입니다. 말은 남의 것이어도, 마구가 좋으면 잘 달립니다.

그래서 추가 비용이 0원입니다

항목	내용
자체 모델	없음
두뇌	GPT · 클로드 등 외부 연결
꽃는 법	이미 쓰고 있는 구독을 그대로 연결
구독이 없으면	API 키를 넣어도 됩니다
추가 비용	0원

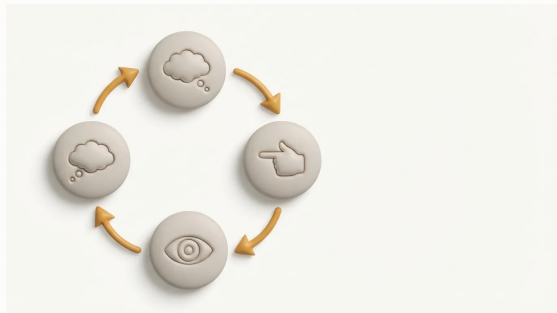
이미 내고 있는 돈으로 쓰는 것이니 새로 나가는 돈이 없습니다. 이게 「무료」의 정체입니다.

두뇌를 직접 만들지 않았다는 것이 약점처럼 들리지만, 오히려 강점입니다. 더 좋은 모델이 나오면 갈아 끼우면 됩니다.

8

루프

생각하고, 행동하고, 본다



사람이 일하는 방식과 같습니다

에이전트는 한 번에 답을 내놓지 않습니다. 네 칸을 계속 돕니다.

생각 → 행동 → 관찰 → 다시 생각 → ...

1

생각

지금 무엇을 해야 하나

2

행동

실제로 눌러 본다

3

관찰

화면이 어떻게 바뀌었나

4

다시 생각

계획을 고친다

계획대로 안 되면 **계획을 고칩니다**. 이 「고침」이 있어서 실제 웹에서 일이 됩니다.

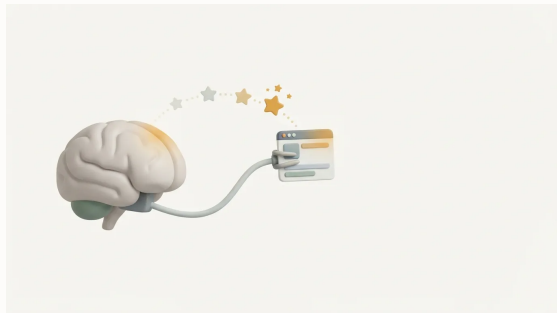
이 루프에 이름이 붙은 게 2022년입니다

이 방식을 정리해 이름을 붙인 논문이 **ReAct** 입니다. Reasoning(추론) + Acting(행동)을 붙인 말입니다.

지금 나오는 거의 모든 에이전트가 이 루프로 돕니다. **도구가 바뀌어도 이 루프는 그대로**입니다. 그래서 이것 하나만 알아도 새 도구를 빨리 이해하실 수 있습니다.

논문 하나 — WebGPT

브라우저를 켜 첫 언어모델



2021년, 오픈AI

항목	내용
기관 · 연도	OpenAI · 2021
정식 제목	WebGPT: Browser-assisted question-answering with human feedback
한 일	GPT-3 에 텍스트 브라우징 환경 을 주어 검색 · 링크 이동 · 인용을 시켰습니다
훈련	사람 시연을 따라 하게 한 뒤, 보상모델을 두고 강화학습
결과	사람 평가 선호 56% (시연자 대비) · 69% (레딧 상위 답변 대비)

arxiv.org/abs/2112.09332

왜 이 논문부터인가

언어모델에게 처음으로 브라우저를 쥐여 준 연구입니다. 지금 보는 모든 브라우저 에이전트의 출발점입니다.

그리고 검색하는 법을 가르친 방법이 **강화학습**이었습니다. AI 가 마우스를 잡는 이야기의 첫 장부터 강화학습이 나옵니다.

잘하면 점수를 주고 못하면 안 주는 식으로 스스로 방법을 찾게 하는 것 — 그게 강화학습입니다. 정답을 글로 적기 어려운 일에 씁니다.

10

논문 둘 — ReAct

루프에 이름을 붙이다

2022년, 프린스턴과 구글

항목	내용
기관 · 연도	프린스턴 · 구글 · 2022
정식 제목	ReAct: Synergizing Reasoning and Acting in Language Models
한 일	추론과 행동을 번갈아 하게 하는 형식을 정립했습니다
효과	계획 수립 · 추적 · 갱신 · 예외 처리 · 외부 정보원 접근

arxiv.org/abs/2210.03629

생각만 하면 헛돌고, 행동만 하면 길을 잃습니다

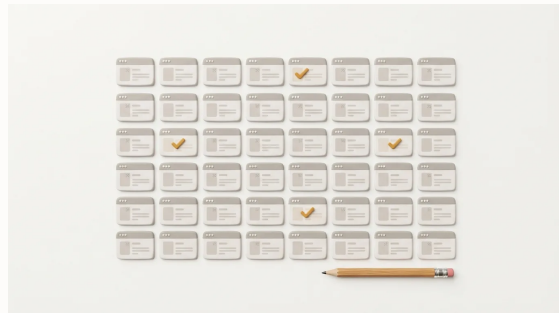
생각만 시키면 그럴듯한 말만 하고 실제로 확인하지 않습니다. 행동만 시키면 왜 그렇게 하는지 모른 채 헤맹니다.

둘을 번갈아 하게 한 것이 이 논문의 기여입니다. 생각 한 번, 행동 한 번, 결과 확인 한 번.

이 형식이 지금 에이전트의 표준이 되었습니다.

논문 셋 — Mind2Web

실제 웹으로 만든 시험지



2023년, 오하이오주립대

항목	내용
기관 · 연도	오하이오주립대 (OSU NLP) · 2023 · NeurIPS Spotlight
정식 제목	Mind2Web: Towards a Generalist Agent for the Web
규모	실제 사이트 137 곳 · 31 개 분야 · 과제 2,000+
특징	모의 환경이 아니라 진짜 웹사이트 . 사람이 실제로 조작한 기록을 모았습니다

arxiv.org/abs/2306.06070

시험지가 왜 중요한가

시험지가 없으면 「우리 것이 제일 낫다」는 말을 검증할 방법이 없습니다.

이 논문이 **공통 시험지**를 만들어 놓은 덕분에, 서로 다른 팀이 만든 에이전트를 같은 기준으로 비교할 수 있게 되었습니다.

다음 장의 「99%」가 바로 이 시험지 점수입니다.

12

99%, 누가 잤나

숫자를 읽는 법



발표된 점수

시험지	점수
Online-Mind2Web	99.0%
BU-Bench-V1	93.0%

세계 최고 수준이라고 발표했습니다. 쓴 두뇌는 GPT 5.5 와 Kimi K2.6 이라고 밝혔습니다.

그런데 채점을 누가 했나

회사가 직접 했습니다. 제3자 검증이 아닙니다.

이건 거짓말이라는 뜻이 아닙니다. 다만 숫자를 볼 때 **항상 같이 봐야** 하는 것이 있습니다.

1

누가 했나

본인인가, 제3자인가

2

무엇으로 했나

공개된 시험지인가, 자기가 만든 것인가

3

방법을 공개했나

따라 해 볼 수 있게 적어 두었는가

이 회사는 **방법을 공개**했습니다. 그건 좋은 신호입니다. 그래도 「자체 측정」이라는 사실은 그대로입니다.

AI 업계 발표를 읽을 때 이 세 가지만 확인하셔도 과장에 덜 흔들리십니다.

13

로컬 AI

완전 무료, 완전 소유



두뇌 자리에 내 컴퓨터를 꽂습니다

7장에서 두뇌를 빌려 온다고 했습니다. 그 자리에 **내 컴퓨터에서 도는 모델**을 꽂을 수 있습니다.

꽂으면	결과
사용료	0원 — 구독도 API 키도 없음
데이터	내 글이 밖으로 나가지 않음
인터넷	두뇌는 내 안에 있음

다만 선이 있습니다

작은 모델로 어디까지 되는지가 갈립니다.

과제	작은 로컬 모델
고르기 · 요약 — 한 단계	잘 됩니다
초안 쓰기 — 여러 단계	아직 버겁습니다

왜 그런가. 단계가 길어질수록 앞에서 무엇을 했는지 기억하면서 판단해야 합니다. 작은 두뇌는 그 기억을 오래 들고 있지 못합니다.

이 선이 「빌려 쓰는 AI 와 소유하는 AI」의 현주소입니다. 작은 두뇌가 커지는 만큼 선은 위로 올라갑니다.

14

안전

임시저장까지만



지켜야 할 세 가지

1

발행하지 마십시오

초안까지만 만들고, **사람이 읽고 고쳐서** 올립니다

2

내 계정으로만

로그인을 상속받아 돕니다. 남의 계정으로 시키지 마십시오

3

처음엔 지켜보십시오

무엇을 누르는지 화면에 보입니다. 몇 번은 옆에서 보십시오

왜 이게 중요한가

자동화가 문제가 되는 건 대개 **사람이 안 보고 발행할 때**입니다.

틀린 정보가 그대로 나가고, 같은 글이 대량으로 올라가고, 플랫폼이 제재합니다. 그러면 자동화를 안 쓴 것만 못합니다.

AI 는 초안까지, 발행은 사람이. 이 선만 지키시면 됩니다.

품질이 곧 신뢰입니다. 빨리 많이 쓰는 것보다 한 편을 제대로 내는 쪽이 길게 갑니다.

15

날말 풀이

이 책에 나온 말들



브라우저와 조종

말	뜻
크로미움	크롬·엣지·웨일이 공통으로 쓰는 오픈소스 뼈대. 구글이 공개했습니다.
포크	오픈소스를 통째로 복제해 자기 방향으로 고쳐 가는 것. 원본은 그대로 남습니다.
접근성 트리	웹페이지의 구조를 글자로 적어 둔 설계도. 원래는 화면 낭독기를 위한 것입니다.
컴퓨터 유즈	화면을 사진으로 보고 좌표를 눌러 조종하는 방식.
Playwright	웹사이트를 코드로 조종하는 문법. 원래 개발자용 시험 도구입니다.
세션 상속	내가 이미 로그인해 둔 상태를 그대로 물려받아 일하는 것.

에이전트와 모델

말	뜻
에이전트	시키는 일만 하는 게 아니라 목표를 받고 스스로 판단해 해내는 AI.
하네스 (마구)	두뇌가 도구를 다루게 해 주는 장치. 두뇌는 남의 것, 마구는 내 것.
ReAct 루프	생각 → 행동 → 관찰 → 다시 생각. 에이전트의 공통 구조.
강화학습	정답을 알려 주는 대신 잘하면 칭찬해 스스로 방법을 찾게 하는 학습법.
벤치마크	여러 AI 를 같은 기준으로 비교하려고 만든 공통 시험지.
로컬 모델	내 컴퓨터 안에서 도는 AI. 사용료가 없고 데이터가 밖으로 안 나갑니다.

이 장은 영상에 나온 설명이 아니라, 읽으실 때 막히지 않도록 따로 정리한 부록입니다.

원문을 직접 보고 싶다면

무엇	주소
WebGPT 논문	arxiv.org/pdf/2112.09332
ReAct 논문	arxiv.org/pdf/2210.03629
Mind2Web 논문	arxiv.org/pdf/2306.06070
크로미움	chromium.org
Playwright	playwright.dev
로컬 AI 도구	lmstudio.ai · ollama.com

전부 무료로 공개되어 있습니다. **남이 요약해 준 것 말고 원문을 한 번 보시는 것이 가장 빠릅니다.**

※ 이 책의 숫자와 사실은 전부 위 원문과 회사 기술 블로그에서 직접 확인했습니다.

다음

챕터 1 · 서울 아파트 두뇌 만들기

새 AI 가 나올 때마다 불안하지 않으려면

챕터 1 읽기

슬라이드 교재는 [여기](#)를 누르세요.