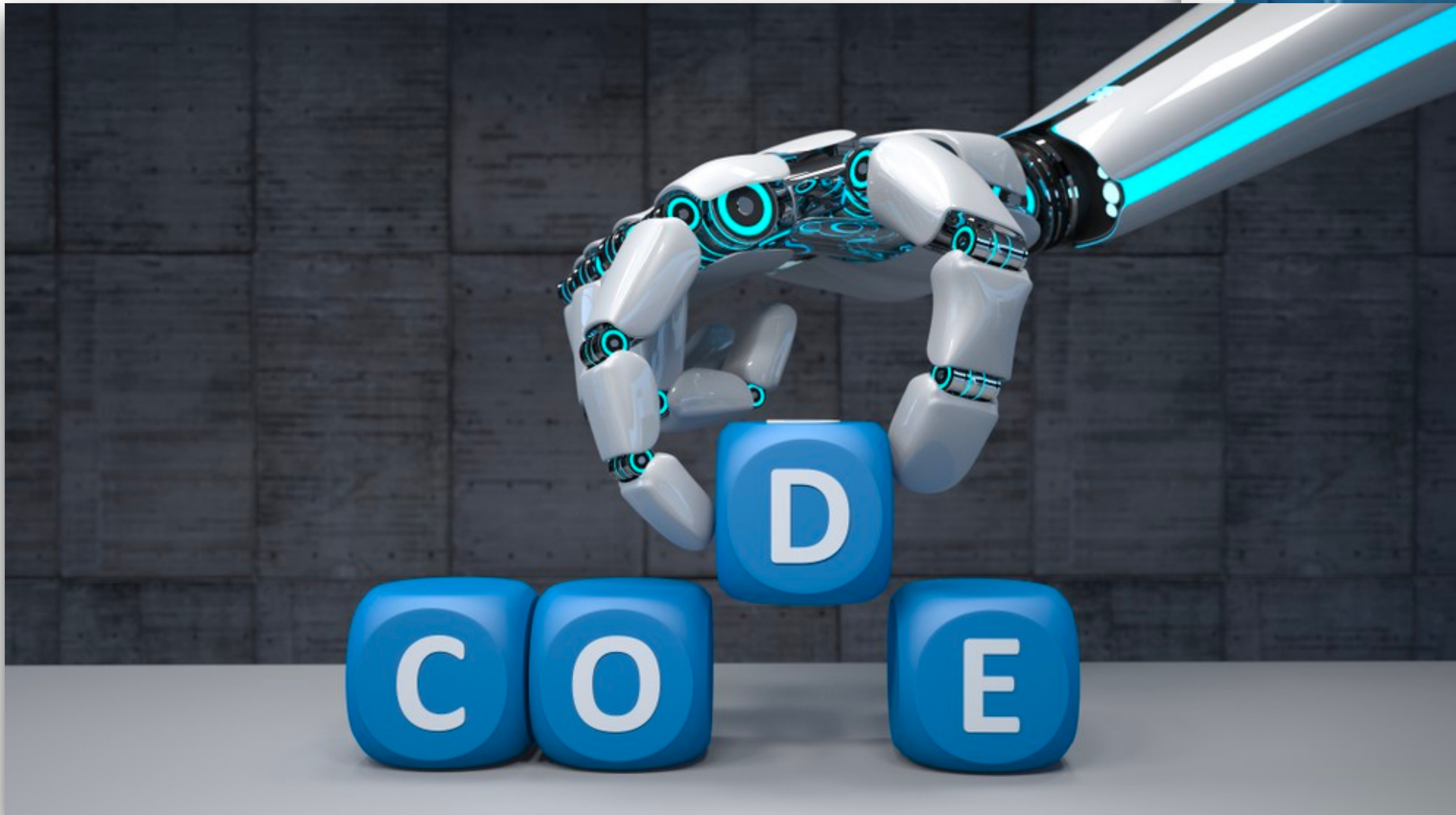


Program Synthesis (Part I)



July 8, 2019

Program Synthesis Moves a Step Closer to Reality

George Leopold



As data scientists and software developers sort through the plethora of tools and APIs ranging from Python to Apache Spark, automation schemes are emerging to help programmers navigate those tools and the accompanying

Apr 16 [Program Synthesis \(Part 1\)](#)
Wed [overview]

Apr 21 [Program Synthesis \(Part 2\)](#)
Mon [overview]

Apr 24 —
Thu

[Exam #2 Due](#)
(you pick a 2-hour window within this
24-hour day)

Apr 25 —
Fri

[HW 6b \(Contribution\) Due](#)

[All Course Materials Due](#)

(cannot be accepted later than this,
please plan accordingly)

Agenda

- A gentle introduction to program synthesis
 - Mostly using examples/demos/figures
 - But also some formalisms/algorithms (Part 1 and Part 2)
- Goals
 - Get exposed to “program synthesis”
 - Appreciate this research topic
 - Learn a little bit about how synthesis works

Example 1

What are we doing here?



The screenshot shows the Microsoft Excel interface. The ribbon at the top includes FILE, HOME, INSERT, PAGE LAYOUT, FORMULAS, and DATA. The active cell is B2, and the formula bar shows 'NL'. The table has two columns: 'Names' (column A) and 'Initials' (column B). The 'Initials' column is split into two sub-columns: 'N L' and 'I'. The data is as follows:

	A	B	
1	Names	Initials	
2	Neil Lieber	N L	I
3	Mathew Prisco		
4	Althea Bertin		
5	Kelly Gamblin		
6	Chandra Valenzula		
7	Cody Castillon		
8	Tyrone Brazier		
9	Althea Buhl		
10	Dollie Munsey		
11	Allyson Phou		

Example 1

What are we doing here?

Filling in Initials for Names



The screenshot shows an Excel spreadsheet with a table. The table has two columns: 'Names' (column A) and 'Initials' (column B). The first row (row 1) contains the headers 'Names' and 'Initials'. The second row (row 2) contains 'Neil Lieber' in column A and 'N L I' in column B. The subsequent rows (rows 3-11) contain names in column A and empty cells in column B. The spreadsheet interface includes the ribbon (FILE, HOME, INSERT, PAGE LAYOUT, FORMULAS, DATA) and the formula bar (B2, X, ✓, ✗, NL).

	A	B
1	Names	Initials
2	Neil Lieber	N L I
3	Mathew Prisco	
4	Althea Bertin	
5	Kelly Gamblin	
6	Chandra Valenzula	
7	Cody Castillon	
8	Tyrone Brazier	
9	Althea Buhl	
10	Dollie Munsey	
11	Allyson Phou	

Example 1

What are we doing here?

Filling in Initials for Names

How would you do it?



The screenshot shows an Excel spreadsheet with a table. The table has two columns: 'Names' (column A) and 'Initials' (column B). The first row (row 1) contains the headers 'Names' and 'Initials'. The second row (row 2) contains 'Neil Lieber' in column A and 'N L I' in column B. The subsequent rows (rows 3-11) contain names in column A and empty cells in column B. The spreadsheet interface includes the ribbon (FILE, HOME, INSERT, PAGE LAYOUT, FORMULAS, DATA) and the formula bar (B2, X, ✓, ✗, NL).

	A	B
1	Names	Initials
2	Neil Lieber	N L I
3	Mathew Prisco	
4	Althea Bertin	
5	Kelly Gamblin	
6	Chandra Valenzula	
7	Cody Castillon	
8	Tyrone Brazier	
9	Althea Buhl	
10	Dollie Munsey	
11	Allyson Phou	

Example 1

What are we doing here?

Filling in Initials for Names

How would you do it?

Idea 1: manually

Idea 2: write an Excel formula?

Idea 3: hire someone to do it for you?

Idea 4: StackOverflow?

Idea 5: ChatGPT?

A screenshot of an Excel spreadsheet. The interface shows the 'FILE' tab selected, with other tabs like 'HOME', 'INSERT', 'PAGE LAYOUT', 'FORMULAS', and 'DATA' visible. The active cell is B2, and the formula bar shows 'NL'. The spreadsheet contains a table with two columns: 'Names' (column A) and 'Initials' (column B). The first row (row 1) contains the headers 'Names' and 'Initials'. The second row (row 2) contains 'Neil Lieber' in column A and 'N L I' in column B. Rows 3 through 11 contain names in column A and empty cells in column B.

	A	B
1	Names	Initials
2	Neil Lieber	N L I
3	Mathew Prisco	
4	Althea Bertin	
5	Kelly Gamblin	
6	Chandra Valenzula	
7	Cody Castillon	
8	Tyrone Brazier	
9	Althea Buhl	
10	Dollie Munsey	
11	Allyson Phou	

Example 1

What are we doing here?

Filling in Initials for Names

How would you do it?

Idea 1: manually

Idea 2: write an Excel formula?

Idea 3: hire someone to do it for you?

Idea 4: StackOverflow?

Idea 5: ChatGPT?

What are some problems with each idea?



The screenshot shows an Excel spreadsheet with a table. The table has two columns: 'Names' (column A) and 'Initials' (column B). The first row (row 1) contains the headers 'Names' and 'Initials'. The second row (row 2) contains 'Neil Lieber' in column A and 'N L I' in column B. The subsequent rows (rows 3-11) contain names in column A and empty cells in column B. The spreadsheet interface includes the ribbon (FILE, HOME, INSERT, PAGE LAYOUT, FORMULAS, DATA) and the formula bar (B2, X, ✓, ✗, NL).

	A	B
1	Names	Initials
2	Neil Lieber	N L I
3	Mathew Prisco	
4	Althea Bertin	
5	Kelly Gamblin	
6	Chandra Valenzula	
7	Cody Castillon	
8	Tyrone Brazier	
9	Althea Buhl	
10	Dollie Munsey	
11	Allyson Phou	

FlashFill

- No (very little) manual work
- No need to write Excel macros
- Highly automated
- People love it



Wesley Backelant
@WesleyBackelant

#FlashFill in #Excel2013 can be so helpful. Love it :-)



Andrew Kemp
@andrew_kemp

Gotta love the flashfill feature in excel 2013 for when I'm working with #PowerShell and #csv for #exchange and #activedirectory

Excel Flash Fill Is A Brilliant Time Saver

Share    |  4

Angus Kidman



Rebecca Jackson 🇯🇵 レベッカ
@RebeccaJLJ

Flashfill in The New Office Excel = witchcraft. It "knows" what you want to do. Love it! - Paul Dolley
#AUSPC



Dr.B 🇺🇸 Learned Hand
@sundarb

Love this back story on the invention of #FlashFill. Well done @SumitGulwani for a really useful tool & to @shobanab Ph.D who was the program manager" why @microsoftexcel is indispensable in B-Schools. #Customer pain points for #Innovation #Marketing

FlashFill Demo

Watch video demo..



What's FlashFill?

- FlashFill is *an Excel feature that automatically transforms strings using examples.*

What's FlashFill?

- FlashFill is an Excel feature that automatically transforms strings using examples.

What does a desired program do in this case?

	A	B
1	Names	Initials
2	Neil Lieber	N L I
3	Mathew Prisco	
4	Althea Bertin	
5	Kelly Gamblin	
6	Chandra Valenzula	
7	Cody Castillon	
8	Tyrone Brazier	
9	Althea Buhl	
10	Dollie Munsey	
11	Allyson Phou	

What's FlashFill?

- FlashFill is an Excel feature that automatically transforms strings using examples.



	A	B
1	Names	Initials
2	Neil Lieber	N L I
3	Mathew Prisco	
4	Althea Bertin	
5	Kelly Gamblin	
6	Chandra Valenzula	
7	Cody Castillon	
8	Tyrone Brazier	
9	Althea Buhl	
10	Dollie Munsey	
11	Allyson Phou	

What does a desired program do in this case?

FlashFill synthesizes a program:

- (1) Take the first char. of first name
- (2) Concatenate it with space
- (3) Concatenate it with the first char. of last name

Guess What FlashFill Would Do?



Guess What FlashFill Would Do?

What (kind of) program do you think FlashFill would synthesize?



Vince Baumel
@quantum_relic

One of my favorite features of #Excel is flash-fill. I can't think of another shortcut that has saved me as much time, and it's surprisingly intuitive. In this example, my string column is actually 3 fields delimited by /. Fill one example row out, and CTRL+E under to flash-fill.

	A	B	C	D
1		NUMBER	NAME	FAVORITE ANIMAL
2	81264/Jessica/Lory, rainbow	81264	Jessica	Lory, rainbow
3	59343/Henryetta/Blue crane			
4	29877/Dacie/Cape raven			
5	30293/Hernando/Macaw, green-winged			
6	78202/Lorenza/Rat, white-faced tree			
7	69532/Laurel/Swan, black			
8	33316/Angelica/Lion, south american sea			
9	23325/Ash/Legaan, ground			
10	76449/Carolyne/Yellow-headed caracara			
11	73305/Urban/Western spotted skunk			
12	45093/Fabian/Shark, blue			
13	26767/Gwenora/Olive baboon			
14	34147/Danna/Boar, wild			
15	57139/Ranee/Violet-crested turaco			
16	31142/Valera/White-tailed deer			
17	72558/Mia/ibex			
18	59246/Deva/Rhinoceros, white			
19	15973/Lorelei/Slender loris			
20	77652/Salomo/Eurasian beaver			
21	67964/Andreas/Argalis			
22	22983/Sherill/Lion, asian			
23	69133/Jordana/European beaver			
24	21496/Averyl/Blue-breasted cordon bleu			
25	46911/Cam/Silver-backed jackal			
26	36133/Bron/Currasow (unidentified)			
27	51294/Candace/Possum, western pygmy			
28	28118/Huntlee/Flightless cormorant			
29	45094/Aguste/Swan, black			
30	27206/Randie/Crowned eagle			
31	79343/Florida/Black-winged stilt			
32	47094/Weston/Leadbeater's ground hornbill			
33	67805/Lionello/Small Indian mongoose			
34	69732/Angelico/Rhinoceros, white			
35	73914/Ker/Blackbird, red-winged			
36	24641/Phaidra/Genet, common			
37	/Amber/Western bearded dragon			
38	27345/Ally/Parrot, hawk-headed			

(next)

Guess What FlashFill Would Do?



Vince Baume
@quantum_relic

One of my favorite features of #Excel is flash-fill. I can't think of another shortcut that has saved me as much time, and it's surprisingly intuitive. In this example, my string column is actually 3 fields delimited by /. Fill one example row out, and CTRL+E under to flash-fill.

What (kind of) program do you think FlashFill would synthesize?

	A	B	C	D
1		NUMBER	NAME	FAVORITE ANIMAL
2	81264/Jessica/Lory, rainbow	81264	Jessica	Lory, rainbow
3	59343/Henryetta/Blue crane			
4	29877/Dacie/Cape raven			
5	30293/Hernando/Macaw, green-winged			
6	78202/Lorenza/Rat, white-faced tree			
7	69532/Laurel/Swan, black			
8	33316/Angelica/Lion, south american sea			

Guess What FlashFill Would Do?

What (kind of) program do you think FlashFill would synthesize?



kurtsh
@kurtsh

I've been working here for more than 2 decades & didn't know this. Damn. [#excel](#) [#powertips](#)

	A	B	C	D	E	F	G
1							
2		Messy Data	First Name	Last Name	Birth Date		
3		Adam Sandylar	4/15/1966	Adam	Sandylar	4/15/1966	
4		Anne Hathahere	1/5/1973				
5		Ariana Venti	12/27/1990				
6		Brad Spitt	2/14/1974				
7		Britney Cheers	4/23/1981				
8		Cardi A	6/19/1977				
9		Christian Ronaldo	12/16/1986				
10		David Backgammon	10/7/1983				
11		Dwayne Jonson	8/23/1989				
12		Ella DeGeneres	6/7/1978				
13		France Hilton	8/13/1977				
14		Gary Busy	3/4/1992				
15		George Looney	7/29/1979				
16		Yennifer Aniston	10/13/1986				
17		John Average	9/13/1971				
18		Johnny Deep	6/21/1978				
19		Justina Bieber	7/16/1982				
20		Katy Berry	4/14/1974				
21		Keanu Leaves	2/19/1976				
22		Kevin Fart	10/6/1965				
23		Khloe Doordashian	4/2/1989				
24		Kim Eyelashian	8/19/1972				
25		Leonardo DiCaprio	1/7/1987				
26		Matt Ramen	3/4/1981				
27		Megan Faux	5/22/1982				
28		Natalie Portapotman	6/22/1965				
29		Nicki Mirage	2/22/1988				

(next)

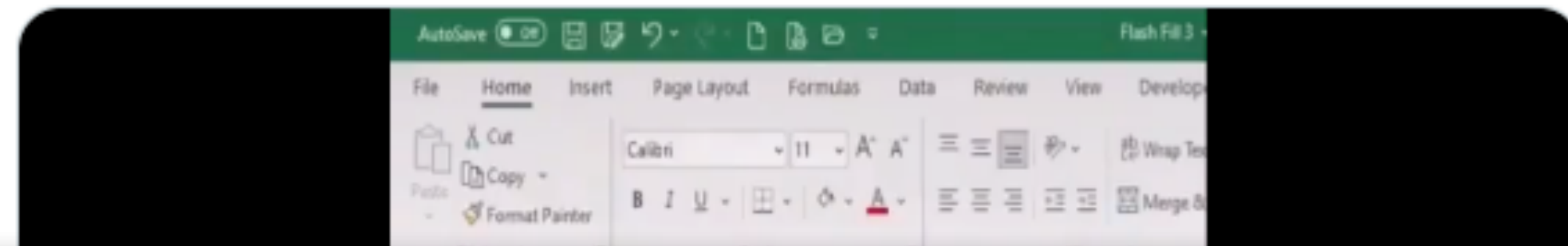
Guess What FlashFill Would Do?

What (kind of) program do you think FlashFill would synthesize?



kurtsh
@kurtsh

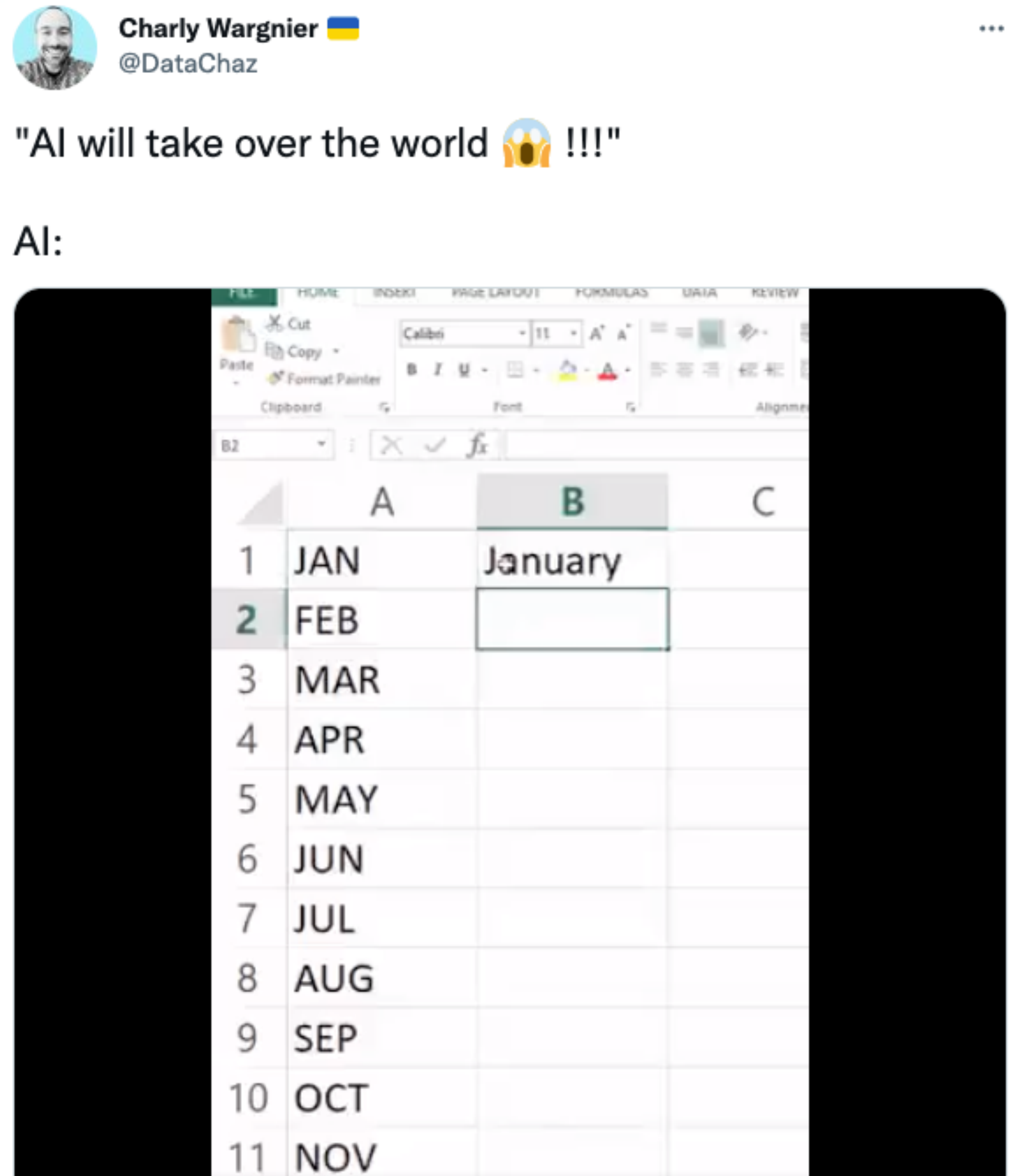
I've been working here for more than 2 decades & didn't know this. Damn. [#excel](#) [#powertips](#)



	A	B	C	D	E	F	G
1							
2		Messy Data	First Name	Last Name	Birth Date		
3		Adam Sand Tyler 4/15/1966	Adam	Sand Tyler	4/15/1966		
4		Anne Hathahere 1/5/1973					
5		Ariana Venti 12/27/1990					
6		Brad Spitt 2/14/1974					

Guess What FlashFill Would Do?

What (kind of) program do you think FlashFill would synthesize?



(no next!)

Guess What FlashFill Would Do?

What (kind of) program do you think FlashFill would synthesize?

Why?



Charly Wagnier 
@DataChaz

"AI will take over the world 🤖 !!!"

AI:

	A	B
1	JAN	January
2	FEB	
3	MAR	
4	APR	
5	MAY	
6	JUN	
7	JUL	
8	AUG	
9	SEP	
10	OCT	
11	NOV	

	A	B
1	JAN	January
2	FEB	Febuary
3	MAR	Maruuary
4	APR	Apruary
5	MAY	Mayuary
6	JUN	Junuary
7	JUL	Juluary
8	AUG	Auguary
9	SEP	Sepuary
10	OCT	Octuary
11	NOV	Novuary

Guess What FlashFill Would Do?

City		No. of
Austin	TX	2
Salt Lake City	UT	1
Durham	NC	3
Columbus	OH	35
Baton Rouge	LA	11
Omaha	NE	27
New Orleans	LA	39
Des Moines	IA	16
Seattle	WA	42
Oklahoma City	OK	14
Houston	TX	620
Charleston	SC	79
Washington	DC	225
Milwaukee	WI	528
Columbia	SC	56
San, Diego	CA	329
Orlando	FL	190
Boston	MA	489
Dallas	TX	489
Minneapolis	MN	396

What (kind of) program do you think FlashFill would synthesize?

(no next!)

Guess What FlashFill Would Do?



City	State	No. of people
Austin	TX	24
Salt Lake City	UT	11
Durham	NC	3
Columbus	OH	35
Baton Rouge	LA	11
Omaha	NE	27
New Orleans	LA	39
Des Moines	IA	16
Seattle	WA	42
Oklahoma City	OK	14
Houston	TX	620
Charleston	SC	70
Washington	DC	225
Milwaukee	WI	528
Columbia	SC	56
San, Diego	CA	329
Orlando	FL	190
Boston	MA	489
Dallas	TX	489
Minneapolis	MN	396

thank you flash fill that is exactly what I wanted

City	State	No. of people
Austin	TX	24
Salt Lake City	UT	11
Durham	NC	3
Columbus	OH	35
Baton Rouge	LA	11
Omaha	NE	27
New Orleans	LA	39
Des Moines	IA	16
Seattle	WA	42
Oklahoma City	OK	14
Houston	TX	620
Charleston	SC	70
Washington	DC	225
Milwaukee	WI	528
Columbia	SC	56
San, Diego	CA	329
Orlando	FL	190
Boston	MA	489
Dallas	TX	489
Minneapolis	MN	396

Why?



FlashFill Doesn't Always Work As Expected...

- FlashFill has an underlying programming language
 - **Domain-specific language (DSL)** for string transformation

FlashFill Doesn't Always Work As Expected...

- FlashFill has an underlying programming language
 - **Domain-specific language (DSL)** for string transformation

Domain-specific language

From Wikipedia, the free encyclopedia

A **domain-specific language (DSL)** is a **computer language** specialized to a particular application **domain**. This is in contrast to a **general-purpose language** (GPL), which is broadly applicable across domains. There are a wide variety of DSLs, ranging from widely used languages for common domains, such as **HTML** for web pages, down to

Some examples: SQL, MATLAB, HTML, etc.

FlashFill Doesn't Always Work As Expected...

- FlashFill has an underlying programming language
 - **Domain-specific language (DSL)** for string transformation
 - .. which means, fundamentally, it can only perform computations expressible in this language
 - In other words, FlashFill cannot synthesize programs that are not expressible in its underlying programming language
 - E.g., transforming dates is not supported, since FlashFill does not recognize dates. Same for US states.

FlashFill T-Shirt!

FlashFill author: Sumit Gulwani (Microsoft)



	A	B	C
1	DEC	December	
2	NOV	November	
3	OCT	October	
4	APR	Aprember	
5	AUG	Augember	
6	FEB	Febember	
7	JAN	Janember	
8	JUL	Julember	
9	JUN	Junember	
10	MAR	Marembere	
11	MAY	Mayember	
12	SEP	Sepember	
13			

Latest Developments of FlashFill

FlashFill++: Scaling Programming by Example by Cutting to the Chase

JOSÉ CAMBRONERO*, Microsoft, USA

SUMIT GULWANI*, Microsoft, USA

VU LE*, Microsoft, USA

FLAME: A Small Language Model for Spreadsheet Formulas

Harshit Joshi^{1*}, Abishai Ebenezer^{4*}, José Cambronero Sanchez^{2†}, Sumit Gulwani^{2†},
Aditya Kanade^{3†}, Vu Le^{2†}, Ivan Radiček^{5†*}, Gust Verbruggen^{2†}

¹ Stanford University

² Microsoft, USA

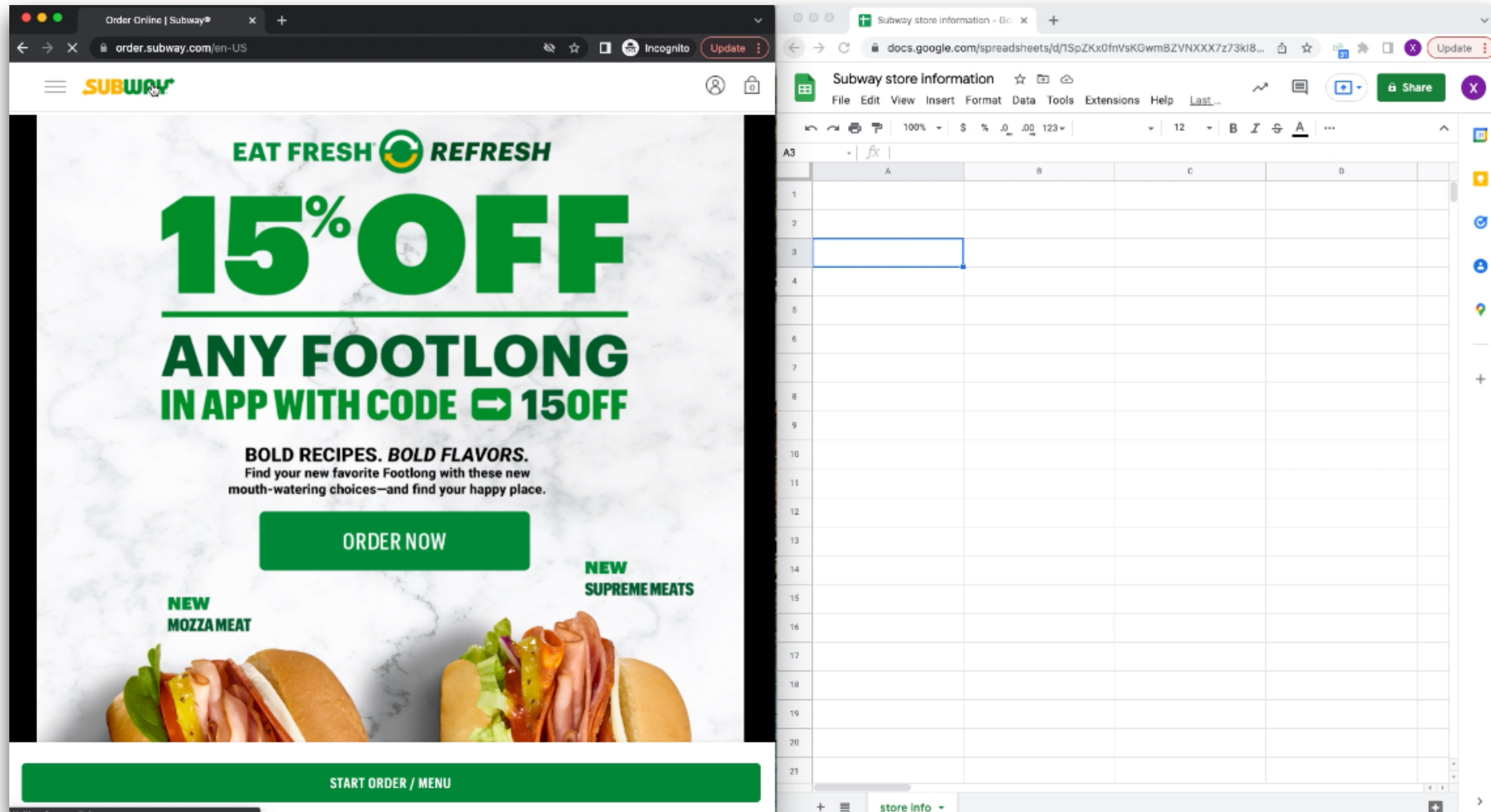
³ Microsoft Research, India

⁴ University of Washington

⁵ Demiurg

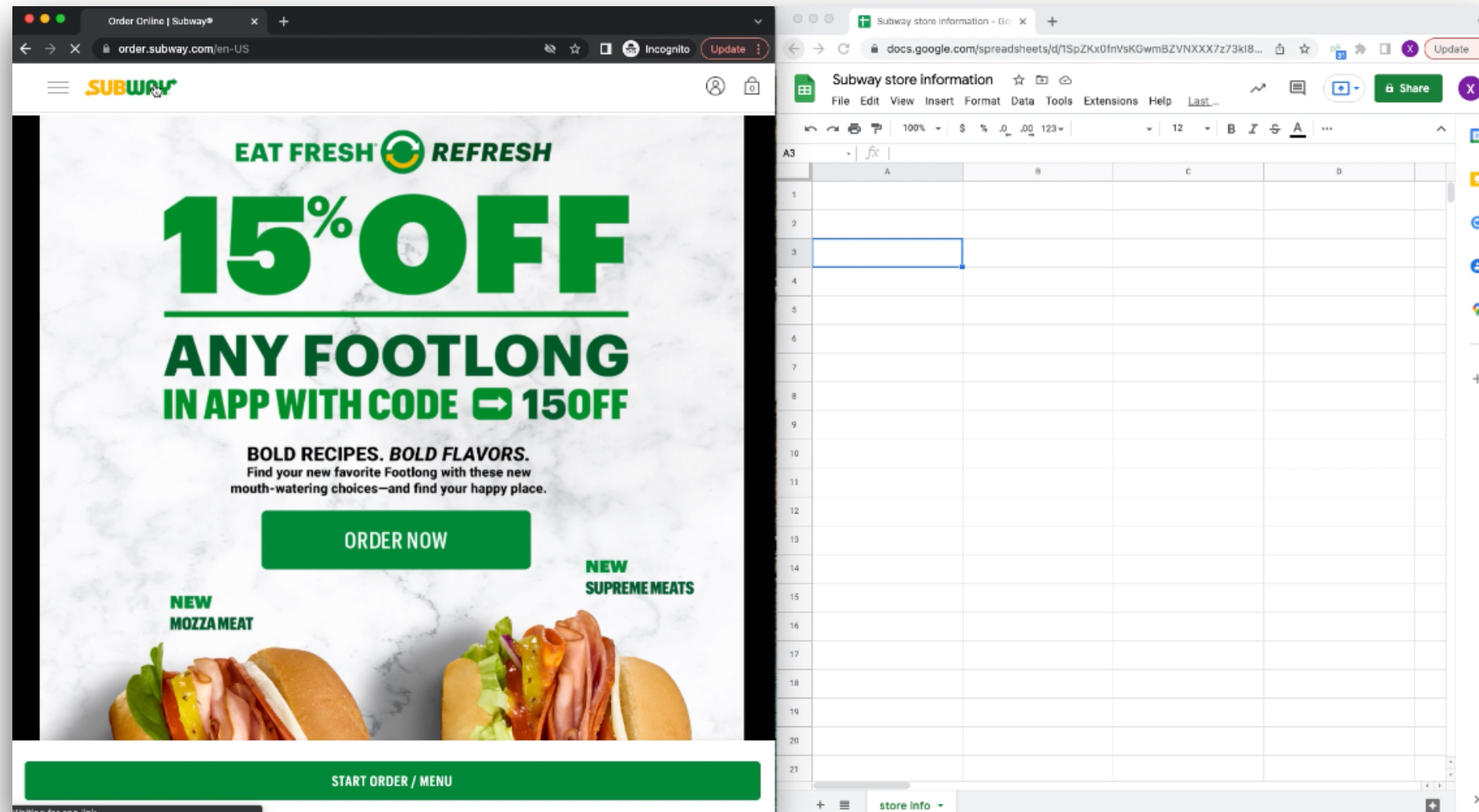
Example 2

Watch video..

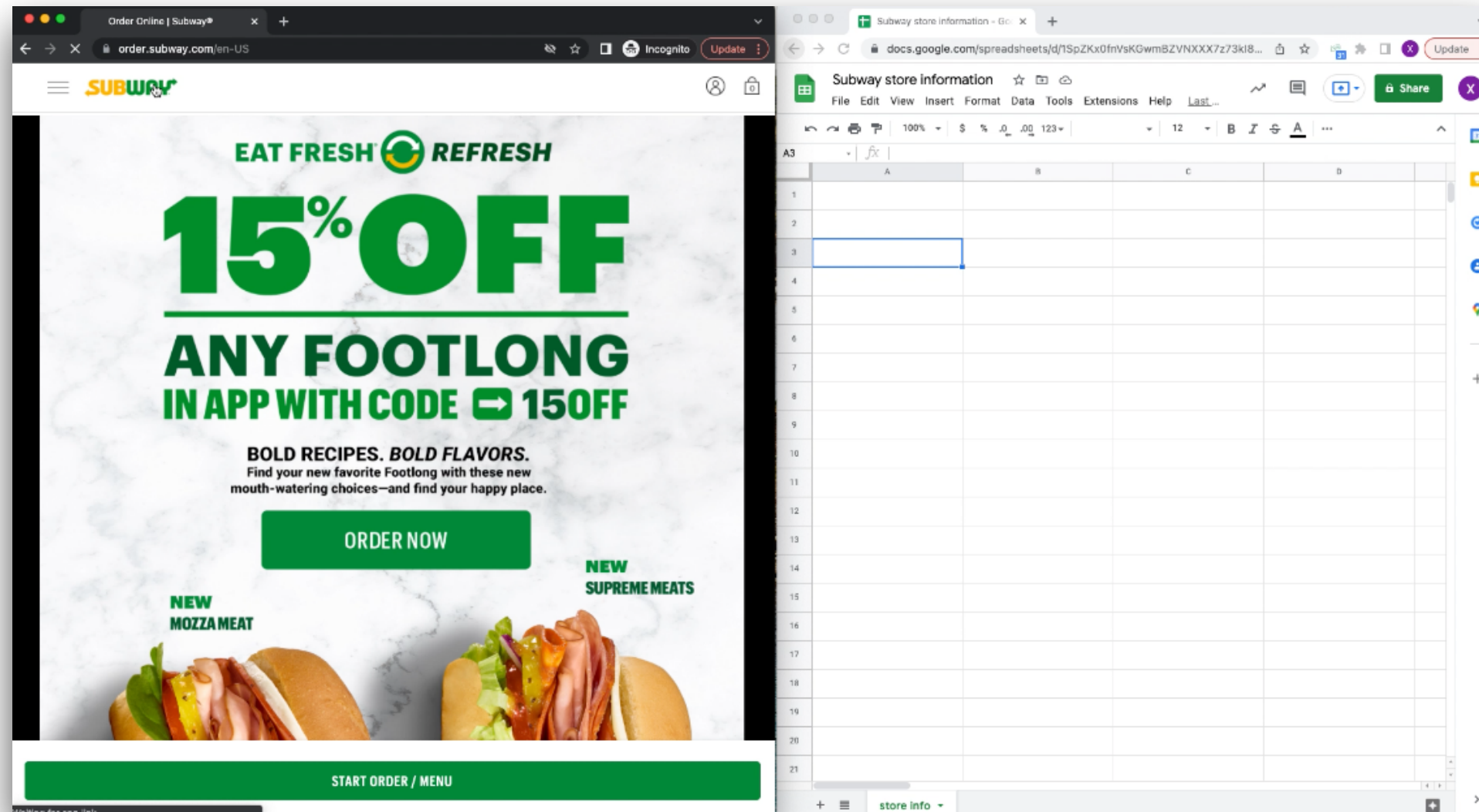


Example 2

What are we doing here?
Scraping Subway store information



Example 2

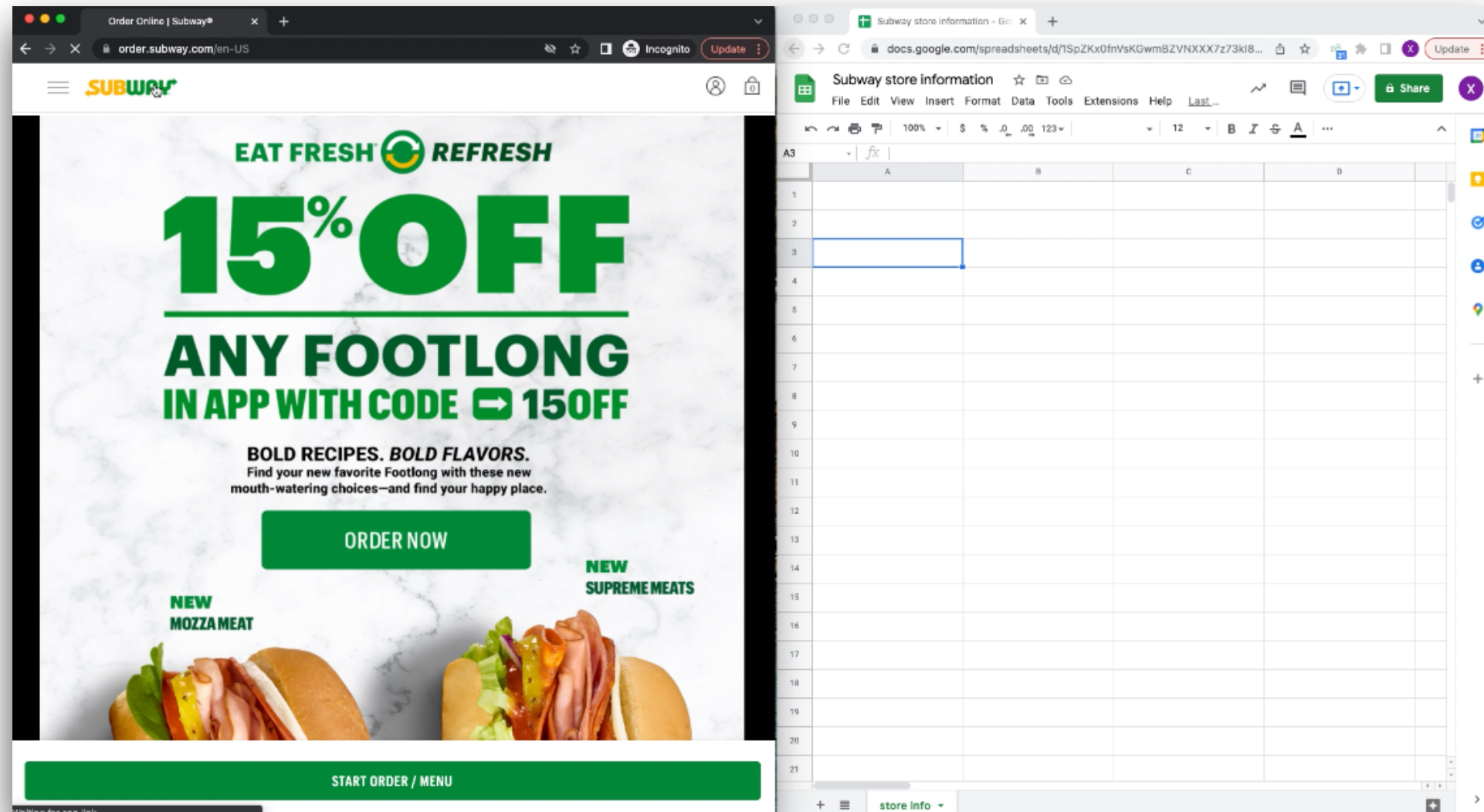


What are we doing here?

Scraping Subway store information

How would you do it?

Example 2



What are we doing here?

Scraping Subway store information

How would you do it?

Idea 1: manually

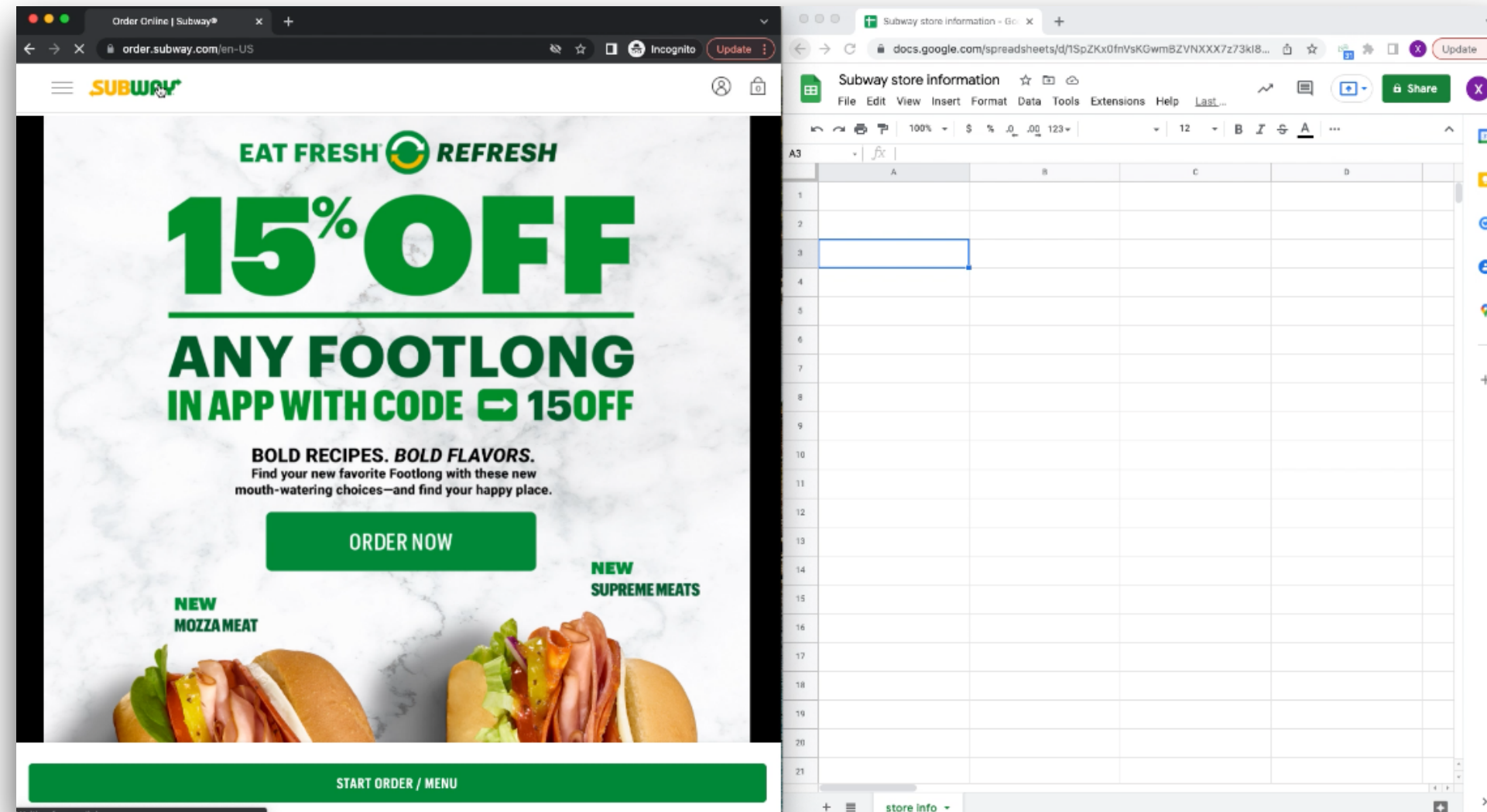
Idea 2: write a Selenium program?

Idea 3: hire someone to do it for you?

Idea 4: StackOverflow?

Idea 5: ChatGPT?

Example 2



What are we doing here?

Scraping Subway store information

How would you do it?

Idea 1: manually

Idea 2: write a Selenium program?

Idea 3: hire someone to do it for you?

Idea 4: StackOverflow?

Idea 5: ChatGPT?

What are some problems with each idea?

WebRobot

- No (very little) manual work
- No need to write Selenium programs
- Highly automated

WebRobot: Web Robotic Process Automation using Interactive Programming-by-Demonstration

Rui Dong

University of Michigan, USA

Zhicheng Huang

University of Michigan, USA

Ian Iong Lam

University of Michigan, USA

Yan Chen

University of Toronto, Canada

Xinyu Wang

University of Michigan, USA



Follow-Ups of WebRobot

WebRobot: Web Robotic Process Automation using Interactive Programming-by-Demonstration

SemanticOn: Specifying Content-Based Semantic Conditions for Web Automation Programs

MIWA: Mixed-Initiative Web Automation for Better User Control and Confidence

DiLogics: Creating Web Automation Programs With Diverse Logics

Efficient Bottom-Up Synthesis for Programs with Local Variables

XIANG LI*, University of Michigan, USA

XIANGYU ZHOU*, University of Michigan, USA

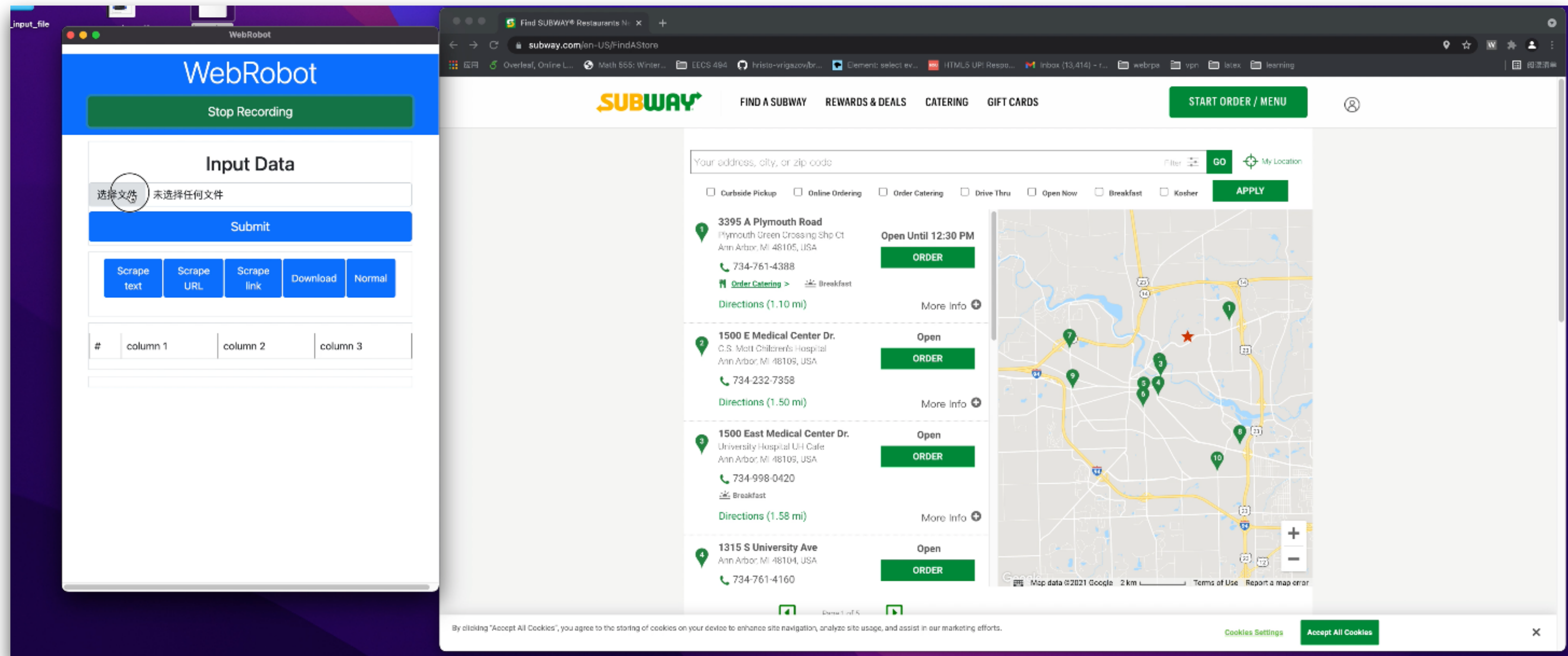
RUI DONG, University of Michigan, USA

YIHONG ZHANG, University of Washington, USA

XINYU WANG, University of Michigan, USA

WebRobot Demo

Watch video demo..



How Does WebRobot Work?

- Idea: Programming-by-Demonstration (PBD)
- Record a **trace** of user-performed actions, then synthesize a **program with loops** that generalizes the trace



```
1 EnterData ../../input x[zip][1]
2 Click ../../button

3 ScrapeText ../../div[1]/div[2]/div[1]/div[1]/div/h3
4 ScrapeText ../../div[1]/div[2]/div[1]/div[2]/div[1]/div[4]/a/div
5 ScrapeText ../../div[2]/div[2]/div[1]/div[1]/div/h3
6 ScrapeText ../../div[2]/div[2]/div[1]/div[2]/div[1]/div[4]/a/div
7 ScrapeText ../../div[3]/div[2]/div[1]/div[1]/div/h3
8 ScrapeText ../../div[3]/div[2]/div[1]/div[2]/div[1]/div[4]/a/div
```

```
EnterData ../../input x[zip][1]
Click ../../button
foreach q in Dscts(e, div[@class='rightContainer']) do
  ScrapeText q//h3
  ScrapeText q//div[@class='locatorPhone']
```

Recap.. FlashFill vs. WebRobot

	FlashFill	WebRobot
Problem	String editing in Excel	Web automation
Idea	Programming-by-Example	Programming-by-Demonstration
DSL	For string editing	For web automation

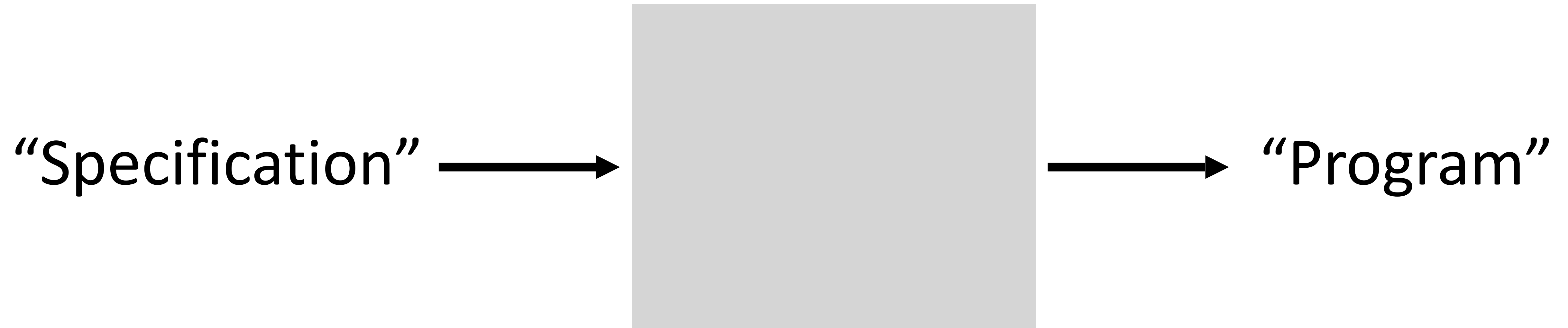
FlashFill vs. WebRobot: What's In Common?

WHAT DO THEY HAVE IN COMMON?



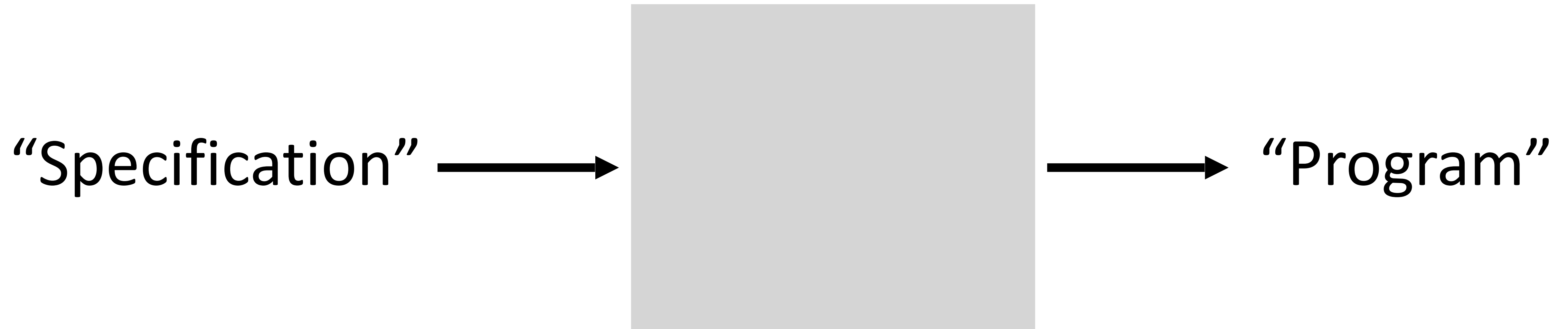
FlashFill vs. WebRobot: What's In Common?

- Share same “interface”.



FlashFill vs. WebRobot: What's In Common?

- Share same “interface”.



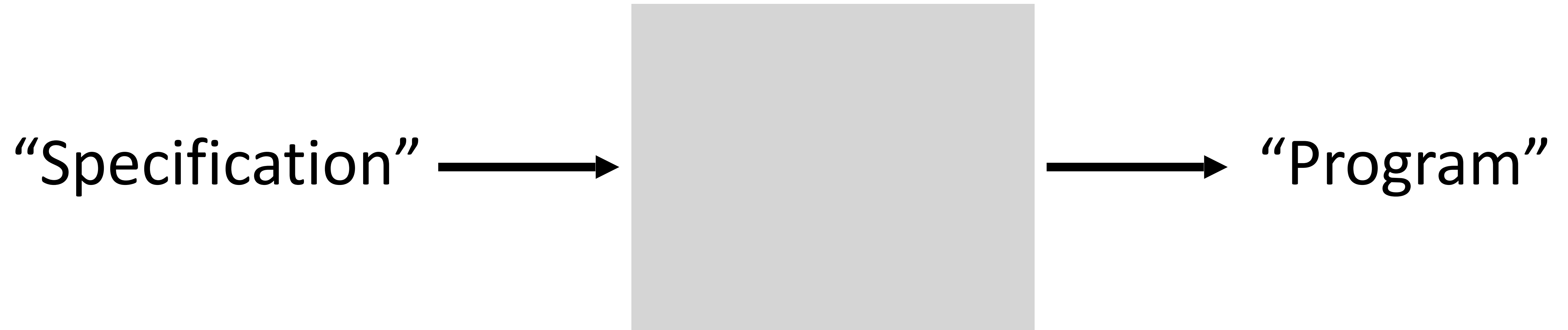
- Differ in:
 - Application domains / use cases.
 - Definitions of “specification”.
 - Programming languages.

“Program Synthesis”



New terminology, but not totally unfamiliar.

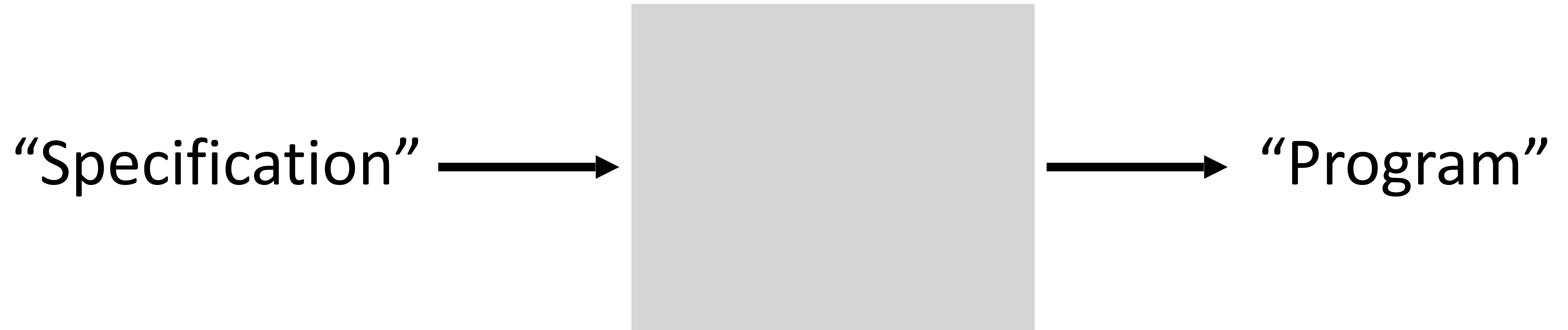
“Program Synthesis”



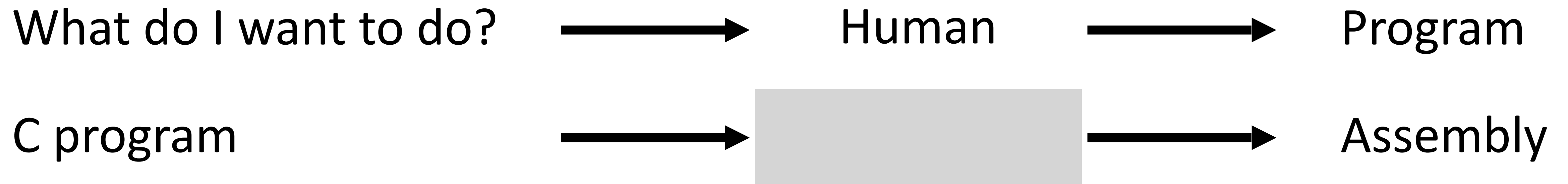
New terminology, but not totally unfamiliar.

What do I want to do?  Human  Program

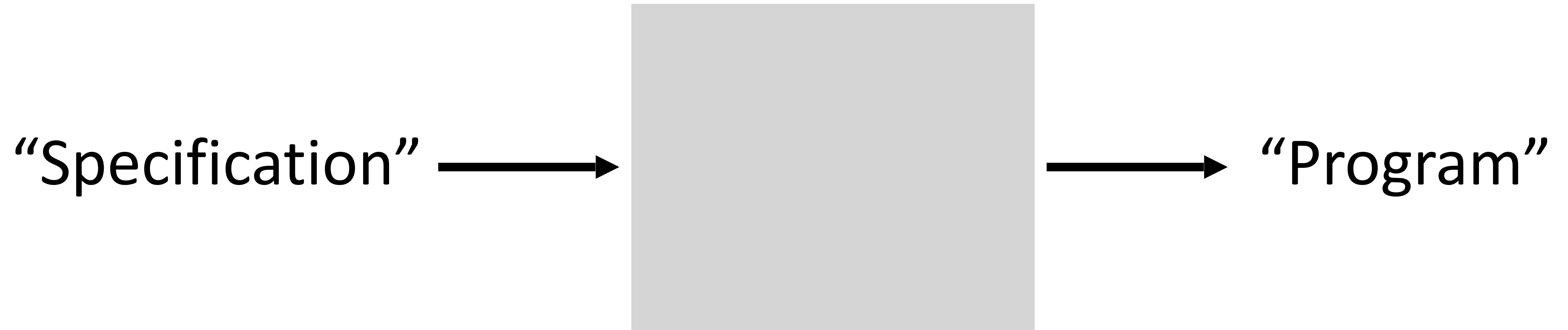
“Program Synthesis”



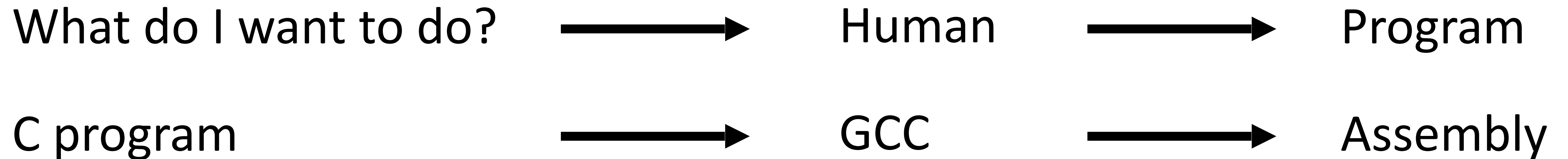
New terminology, but not totally unfamiliar.



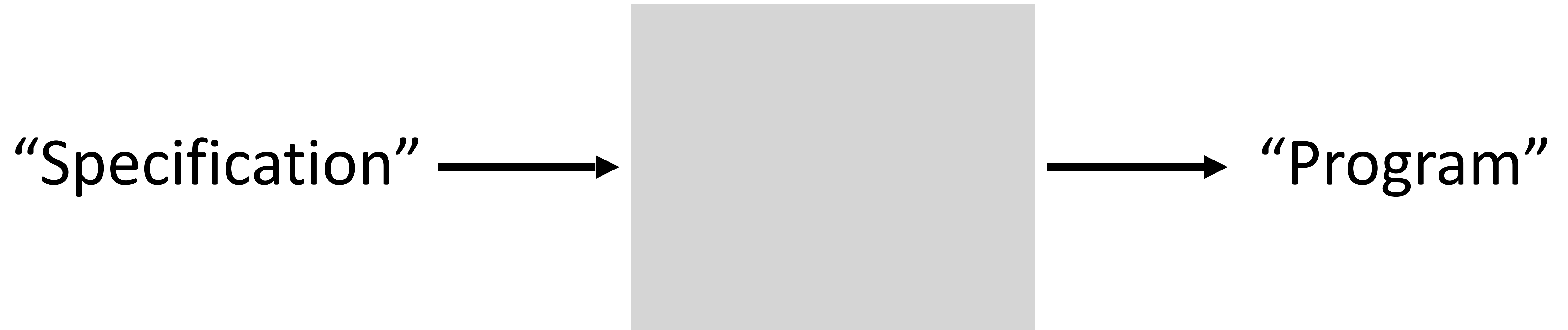
“Program Synthesis”



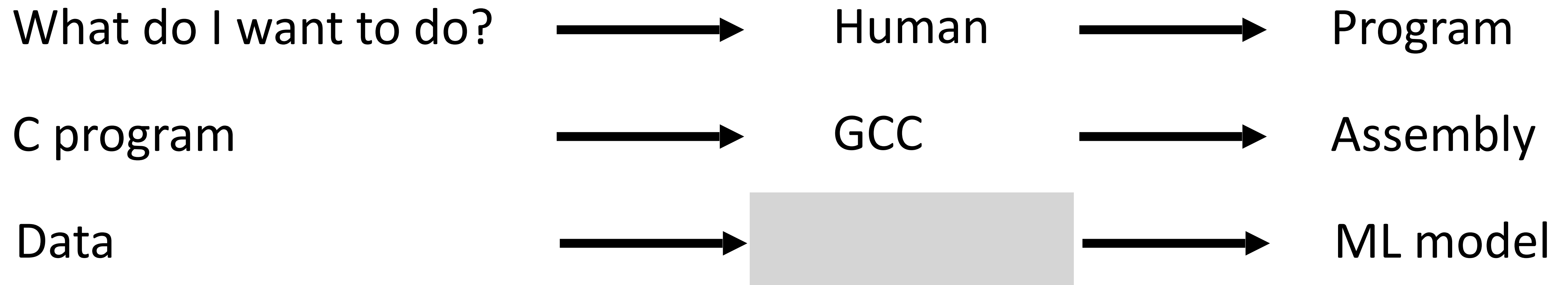
New terminology, but not totally unfamiliar.



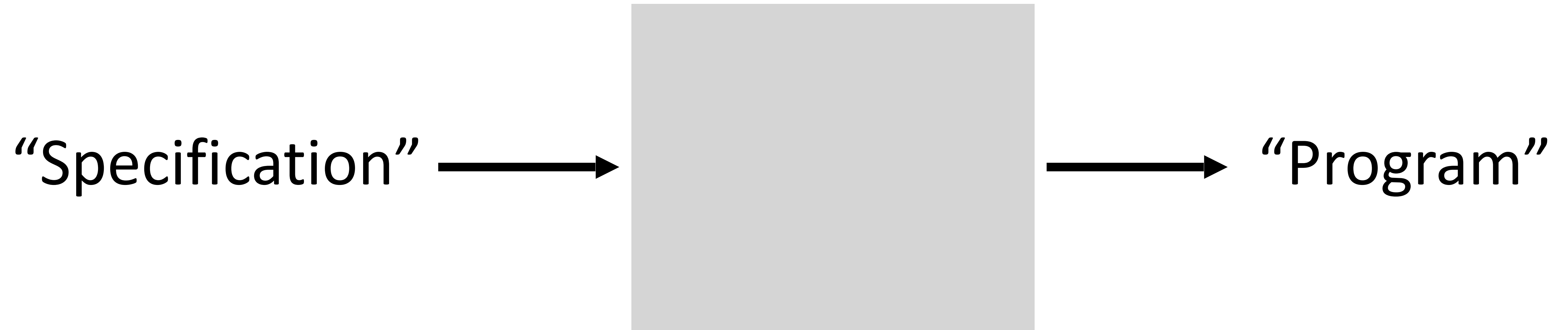
“Program Synthesis”



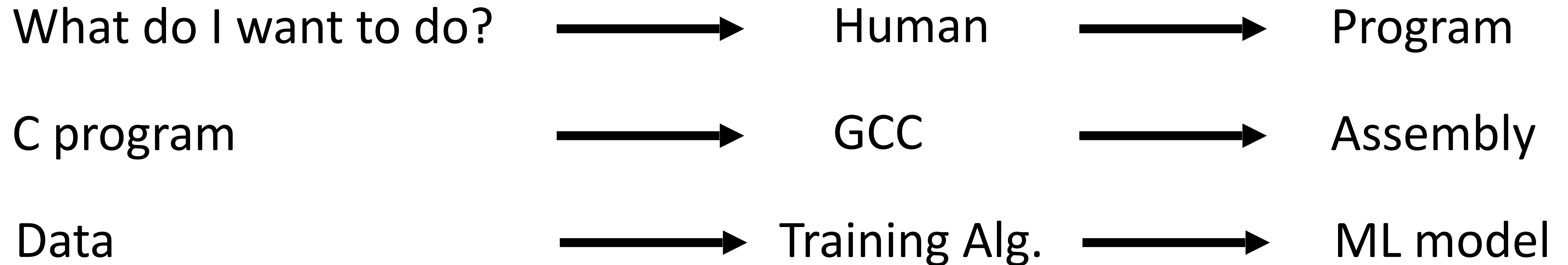
New terminology, but not totally unfamiliar.



“Program Synthesis”

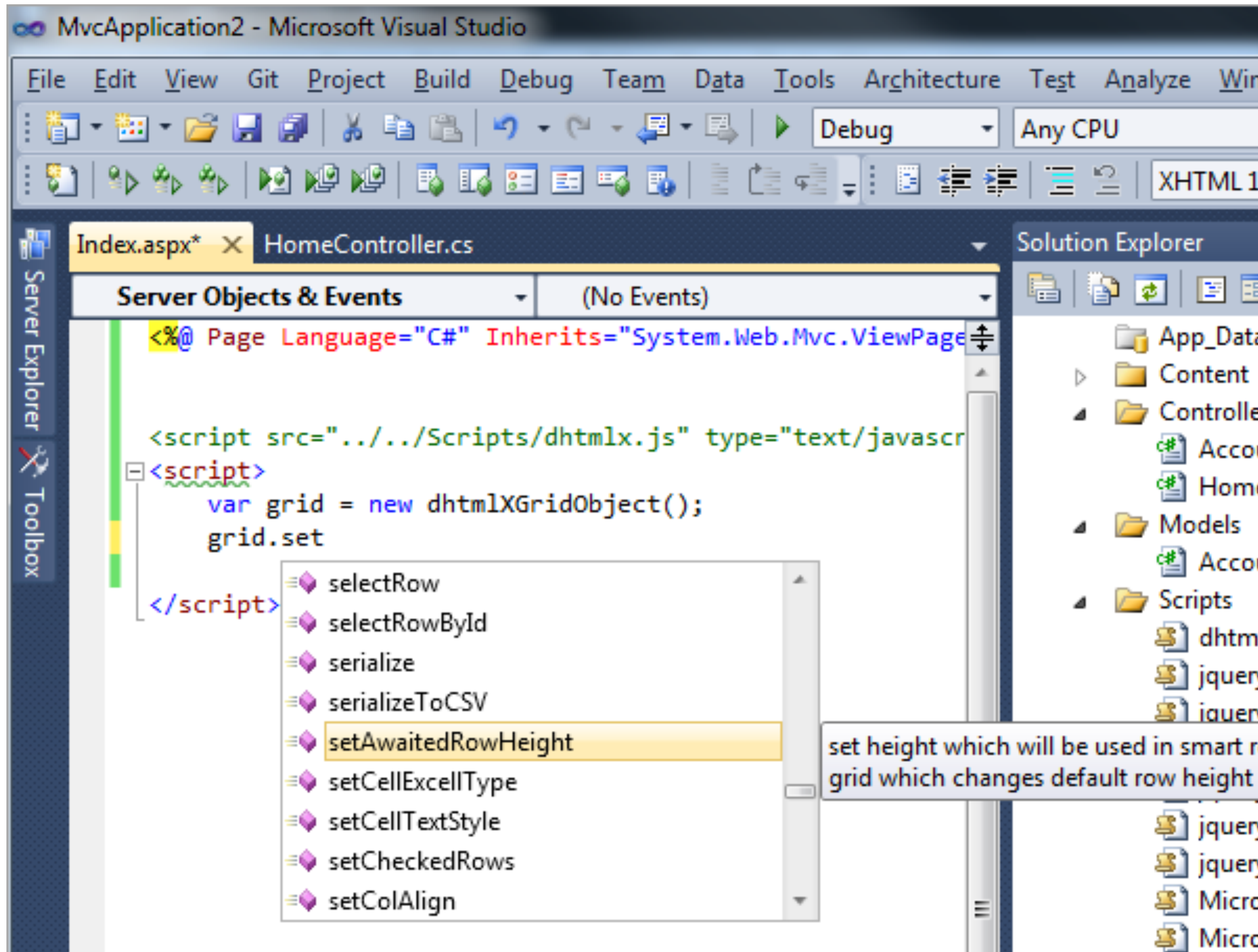


New terminology, but not totally unfamiliar.



“Program Synthesis”

How about this?



“Program Synthesis”

How about this?

	A	B	C
1	Name and ID	First name and last name	ID #
2	Thomas, Rhonda 82132	Rhonda Thomas	
3	Emmett, Keara 34231	Keara Emmett	
4	Vogel, James 32493	James Vogel	
5	Jelen, Bill 23911	Bill Jelen	
6	Miller, Sylvia 78356	Sylvia Miller	
7	Lambert, Bobby 25900	Bobby Lambert	
8	Sweet, Julie 65477	Julie Sweet	
9	Williams, Don 43920	Don Williams	
10	Spake, Deborah 33488	Deborah Spake	



“Program Synthesis”

“Specification” →  → “Program”

What do I want to do?	→	Human	→	Program
C program	→	GCC	→	Assembly
Data	→	Training Alg.	→	ML model
Partial prog.	→	Autocomplete	→	Complete prog.
Examples	→	FlashFill	→	FlashFill prog.
Demonstrations	→	WebRobot	→	Selenium prog.

Working Definition of Program Synthesis

High-level intent

Specification

Program synthesis



Lower-level code

Program

Typically involves search

*I/O examples, demonstrations,
natural language, reference
implementation, etc.*

*In some programming language
(grammar + semantics)*

Program Synthesis vs. Machine Learning/Deep Learning

- Is ML/DL also program synthesis?

Program Synthesis vs. Machine Learning/Deep Learning

- Is ML/DL also program synthesis?
 - ML/DL: data is spec, model is program, try to learn a model that matches data
 - At a high-level, yes

Program Synthesis vs. Machine Learning/Deep Learning

- Is ML/DL also program synthesis?
 - ML/DL: data is spec, model is program, try to learn a model that matches data
 - At a high-level, yes
 - But not really..
 - Definitions of “programs” are very different (e.g., grammar vs. neural nets)
 - Data is noisy but spec is less noisy (but there is a trend in program synthesis to tolerate noise in spec)
 - Typically continuous in ML/DL vs. discrete search space in program synthesis
 - The line is getting blurry

Program Synthesis vs. Compilers

- Are program synthesizers compilers? Are compilers synthesizers?

Program Synthesis vs. Compilers

- Are program synthesizers compilers? Are compilers synthesizers?
 - Compilers also convert high-level intent (code) to lower-level code
 - At a high-level, yes

Program Synthesis vs. Compilers

- Are program synthesizers compilers? Are compilers synthesizers?
 - Compilers also convert high-level intent (code) to lower-level code
 - At a high-level, yes
 - But not really..
 - Compilers translate (not really true anymore) but synthesizers discover
 - Compilers apply predefined transformations (again, not really true nowadays) whereas synthesizers perform search
 - The line is getting blurry

Program Synthesis vs. Compilers

- Are program synthesizers compilers? Are compilers synthesizers?



Sam Tobin-Hochstadt

@samth

...

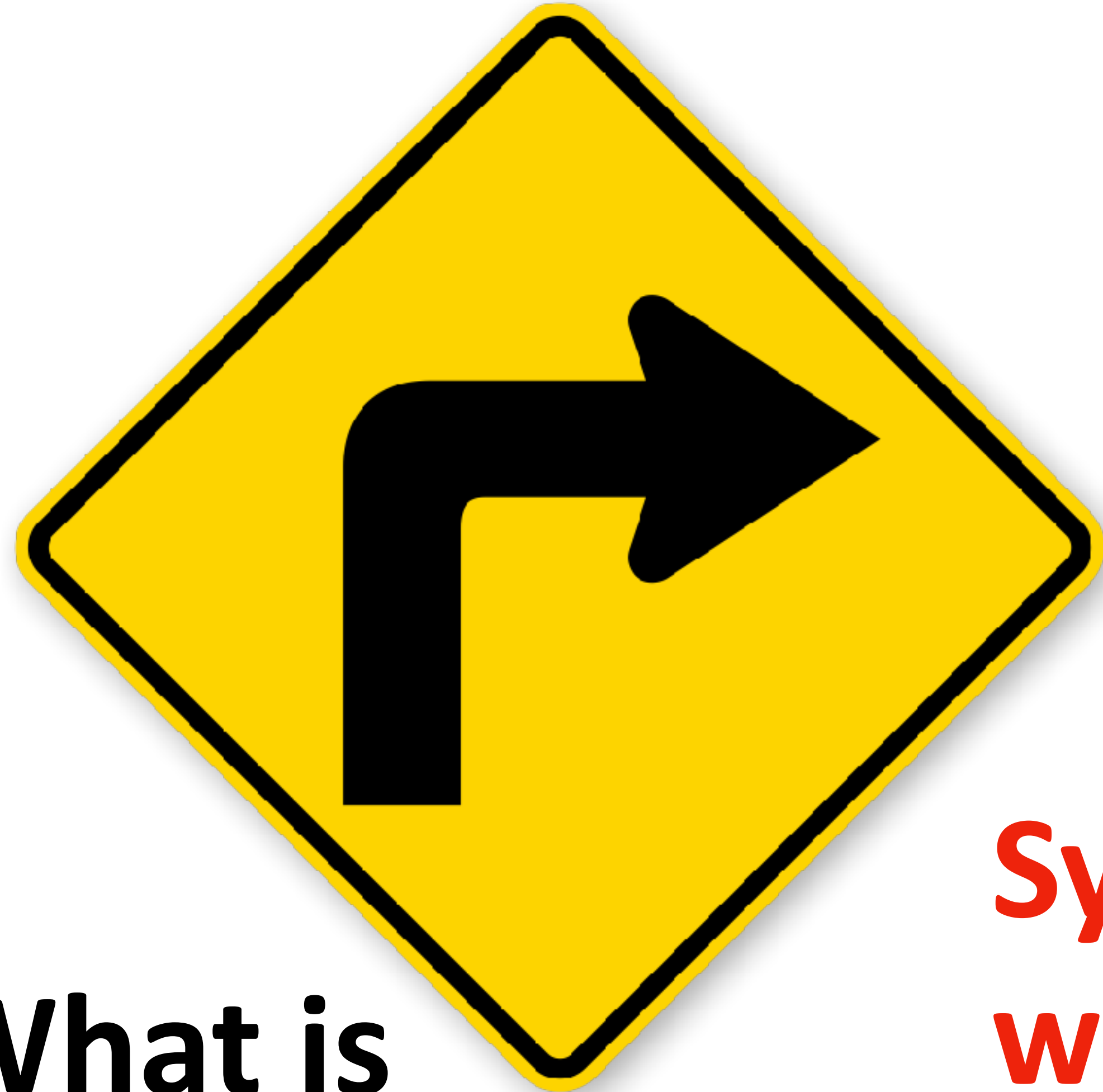
Replying to [@notypes](#)

I was once at a synthesis talk in which David Grove quoted Eran Yahav(?) as saying that a synthesizer is just a compiler that doesn't work. I found that definition really helpful, even if it was sort of a joke.

3:20 PM · May 6, 2020 · Twitter for Android

ver

Switch Gears..



How to
synthesize?



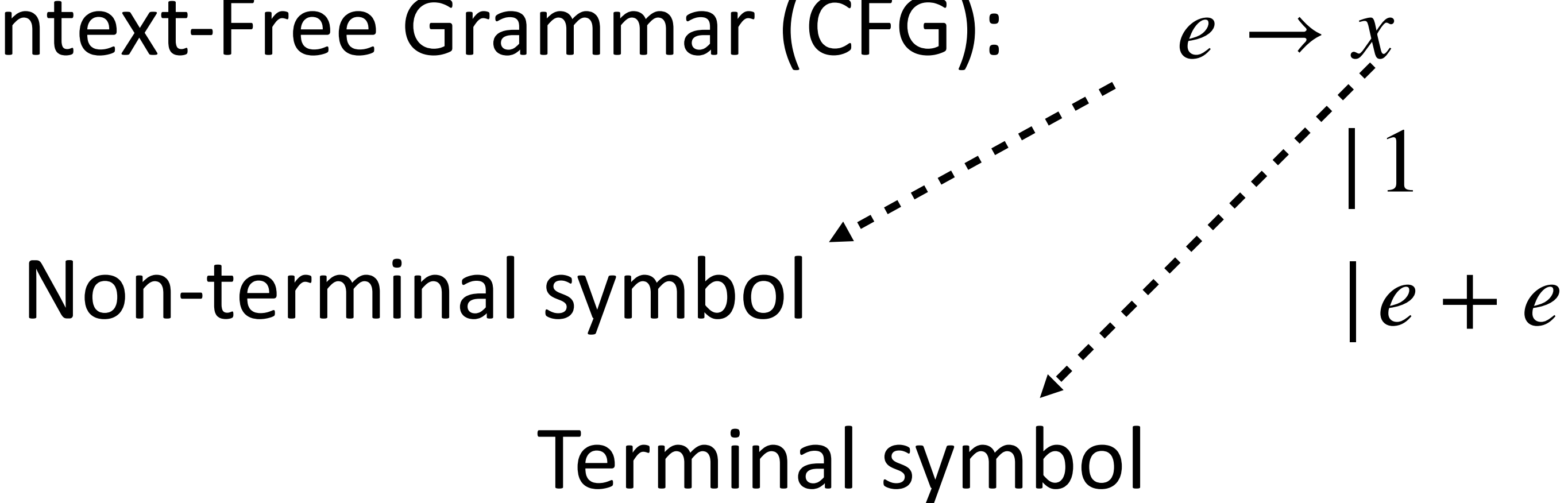
Math!

What is
synthesis?

**Synthesis algorithms
will NOT be covered in
Exam 2**

Systematically Search Programs in Grammar

- Context-Free Grammar (CFG):

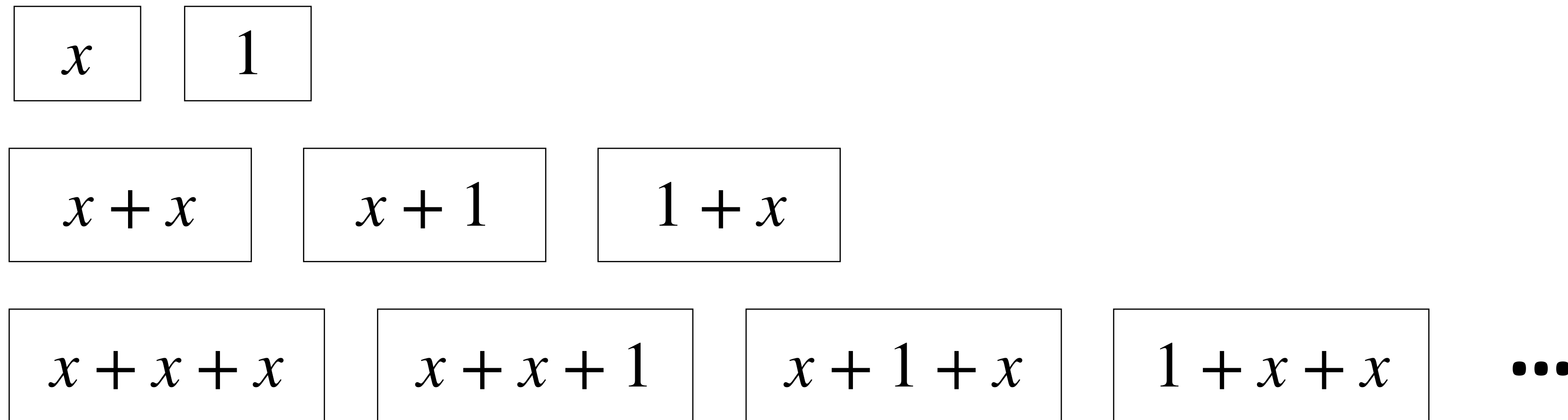


“All programs that can use x , 1 and $+$.”

Systematically Search Programs in Grammar

- Context-Free Grammar (CFG):
$$\begin{array}{l} e \rightarrow x \\ \quad | 1 \\ \quad | e + e \end{array}$$
“All programs that can use x , 1 and $+$.”

- This CFG defines a set of programs

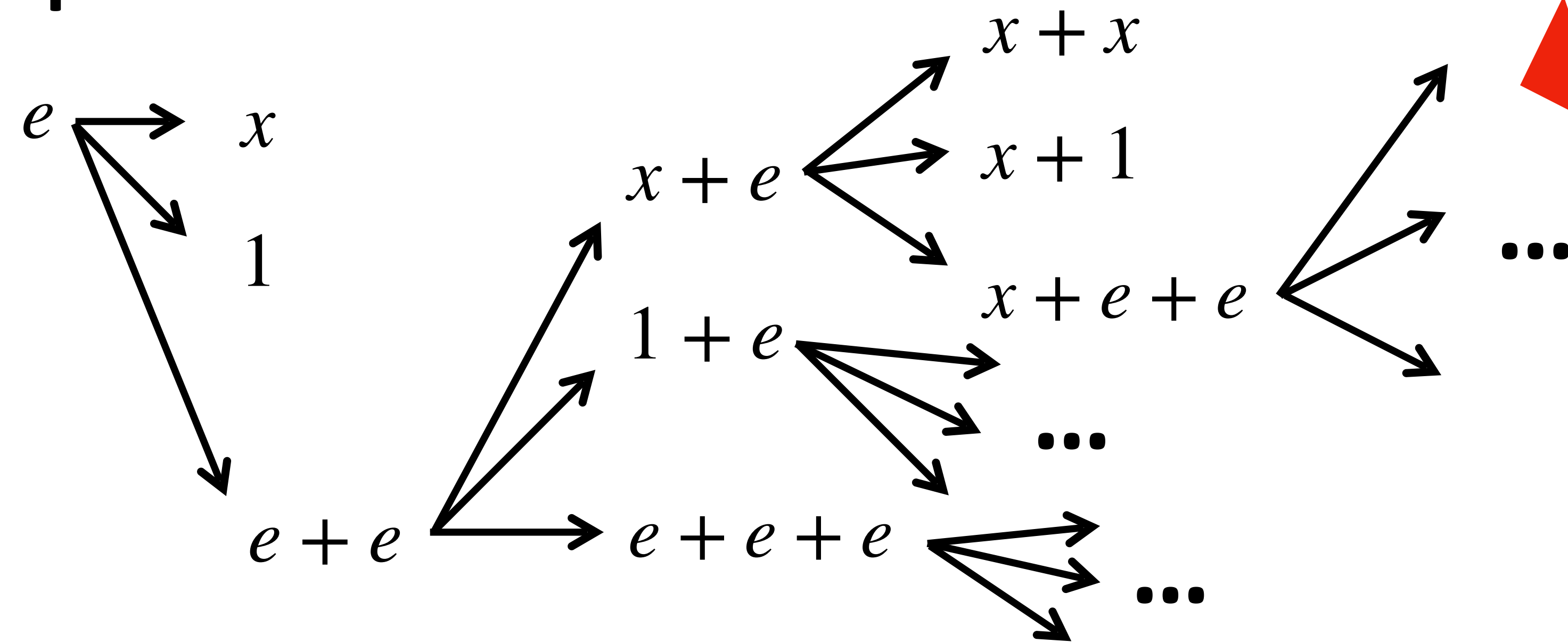


- Goal: find a program in this set that satisfies a given example

How to Search Systematically?

- Context-Free Grammar (CFG):
$$\begin{array}{l} e \rightarrow x \\ \quad | 1 \\ \quad | e + e \end{array}$$

- Top-down search**



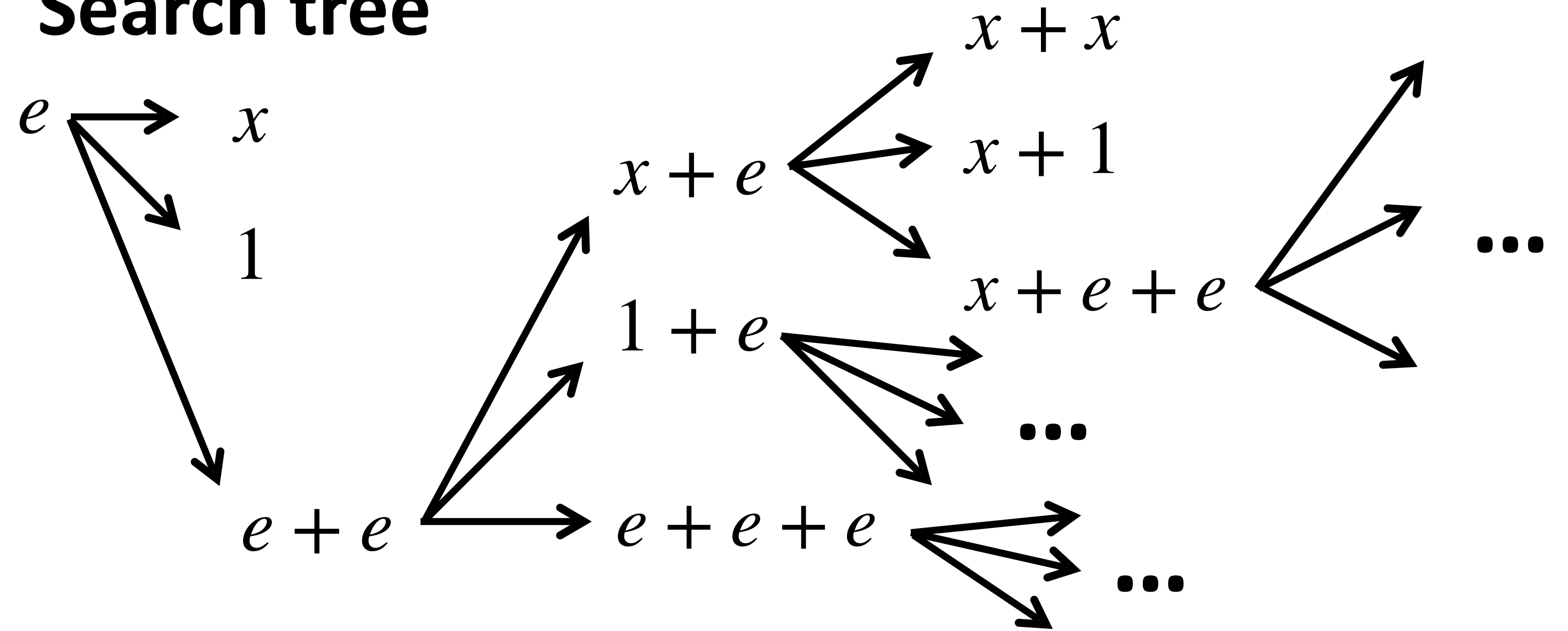
This is a “search tree”

Top-Down Search In Action

Context-Free Grammar

$$\begin{array}{l} e \rightarrow x \\ \quad | 1 \\ \quad | e + e \end{array}$$

Search tree



How to generate this search tree?

- Step 1: begin with the start symbol
- Step 2: pick a non-terminal in current result and replace it with one of its productions
- Step 3: continue step 2 until no more non-terminal remains (i.e., only terminals)

Top-Down Search In Action

Context-Free Grammar

$e \rightarrow x$

| 1

| $e + e$

Search tree

e

How to generate this search tree?

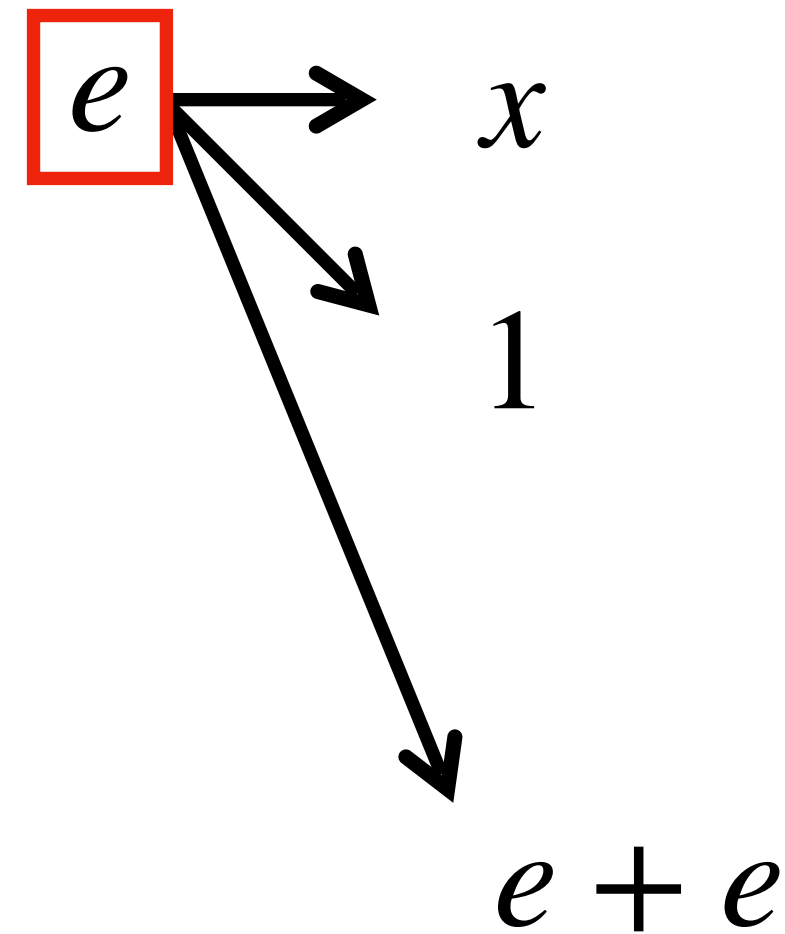
- **Step 1: begin with the start symbol**
- Step 2: pick a non-terminal in current result and replace it with one of its productions
- Step 3: continue step 2 until no more non-terminal remains (i.e., only terminals)

Top-Down Search In Action

Context-Free Grammar

$e \rightarrow x$
| 1
| $e + e$

Search tree



How to generate this search tree?

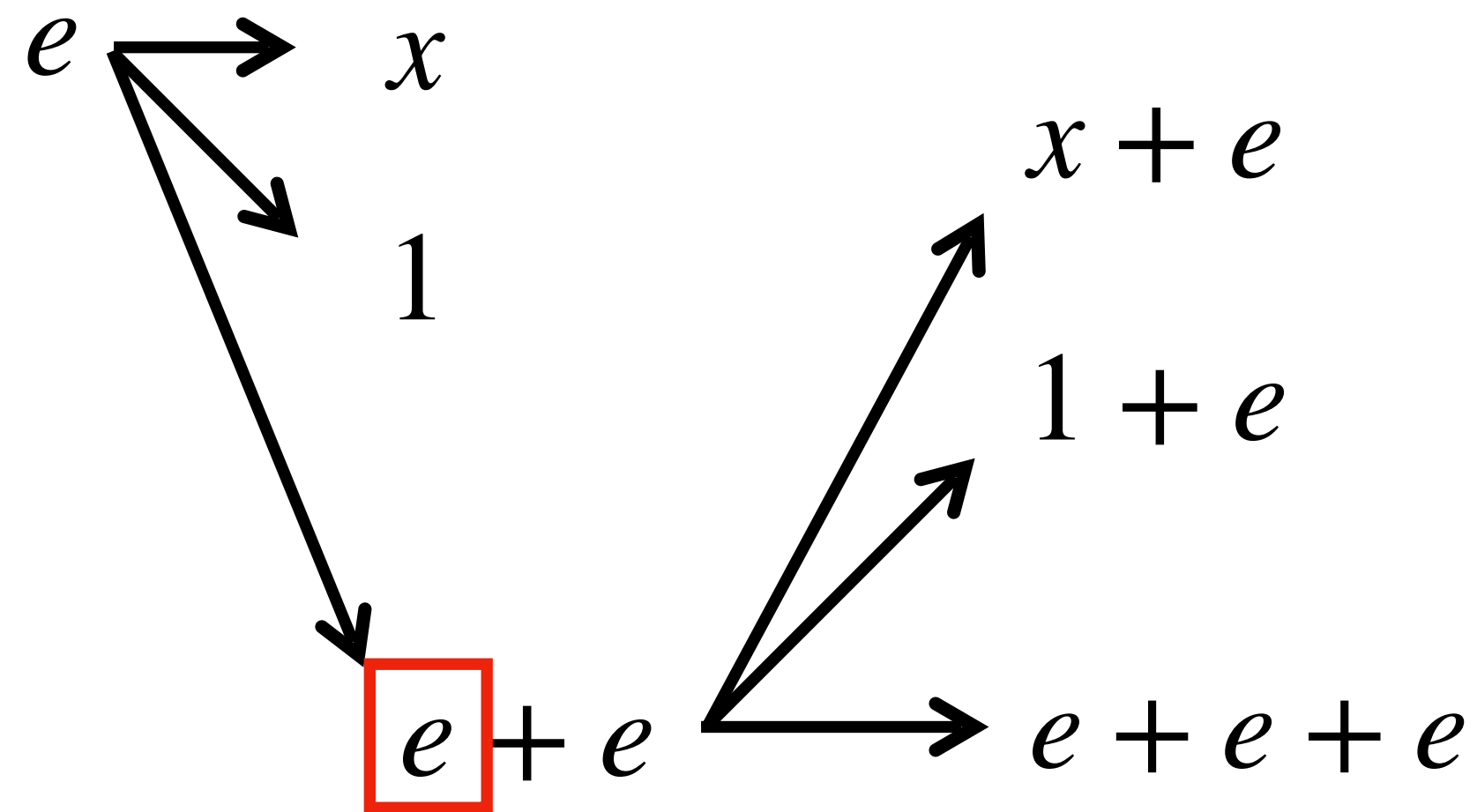
- Step 1: begin with the start symbol
- **Step 2: pick a non-terminal in current result and replace it with one of its productions**
- Step 3: continue step 2 until no more non-terminal remains (i.e., only terminals)

Top-Down Search In Action

Context-Free Grammar

$$\begin{array}{l} e \rightarrow x \\ \quad | 1 \\ \quad | e + e \end{array}$$

Search tree



How to generate this search tree?

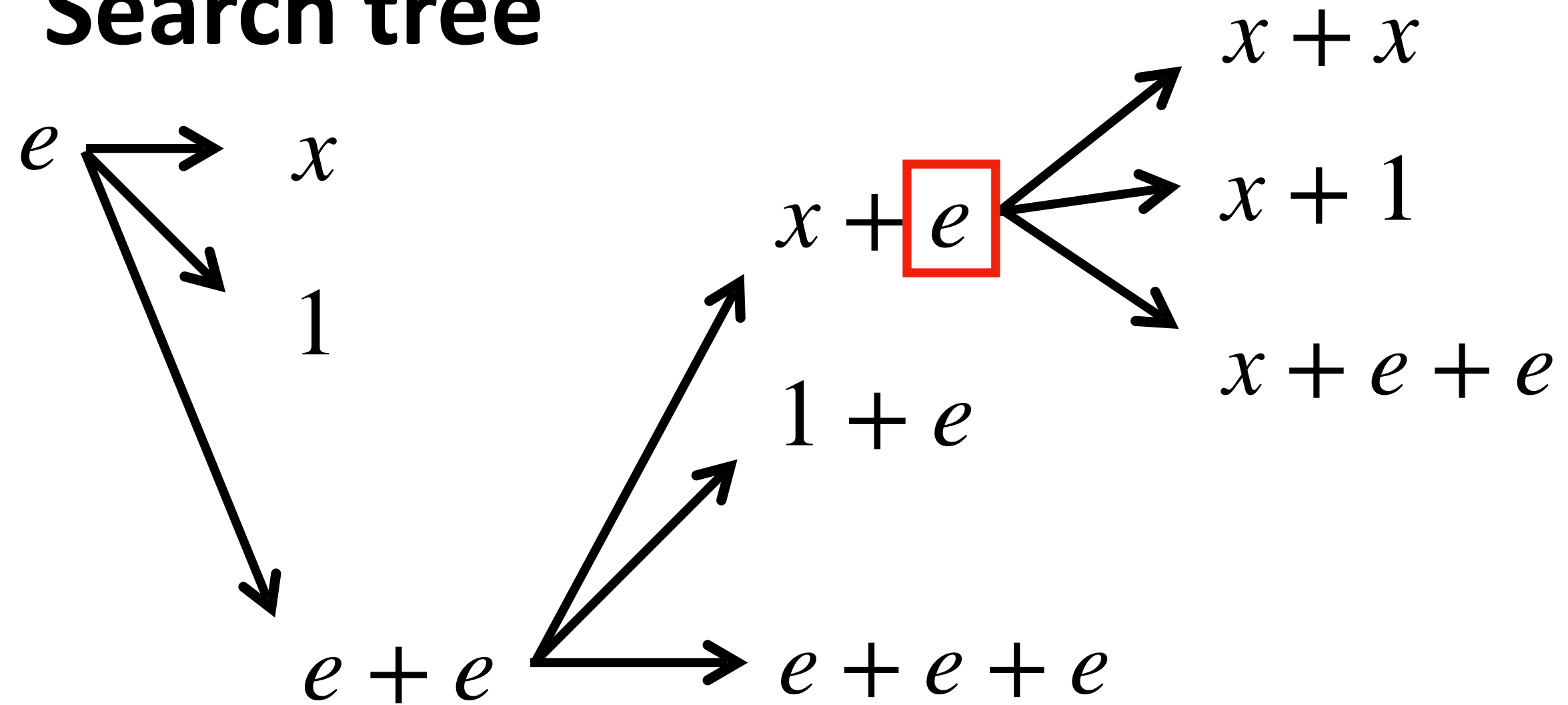
- Step 1: begin with the start symbol
- Step 2: pick a non-terminal in current result and replace it with one of its productions
- **Step 3: continue step 2 until no more non-terminal remains (i.e., only terminals)**

Top-Down Search In Action

Context-Free Grammar

$$\begin{array}{l} e \rightarrow x \\ \quad | 1 \\ \quad | e + e \end{array}$$

Search tree



How to generate this search tree?

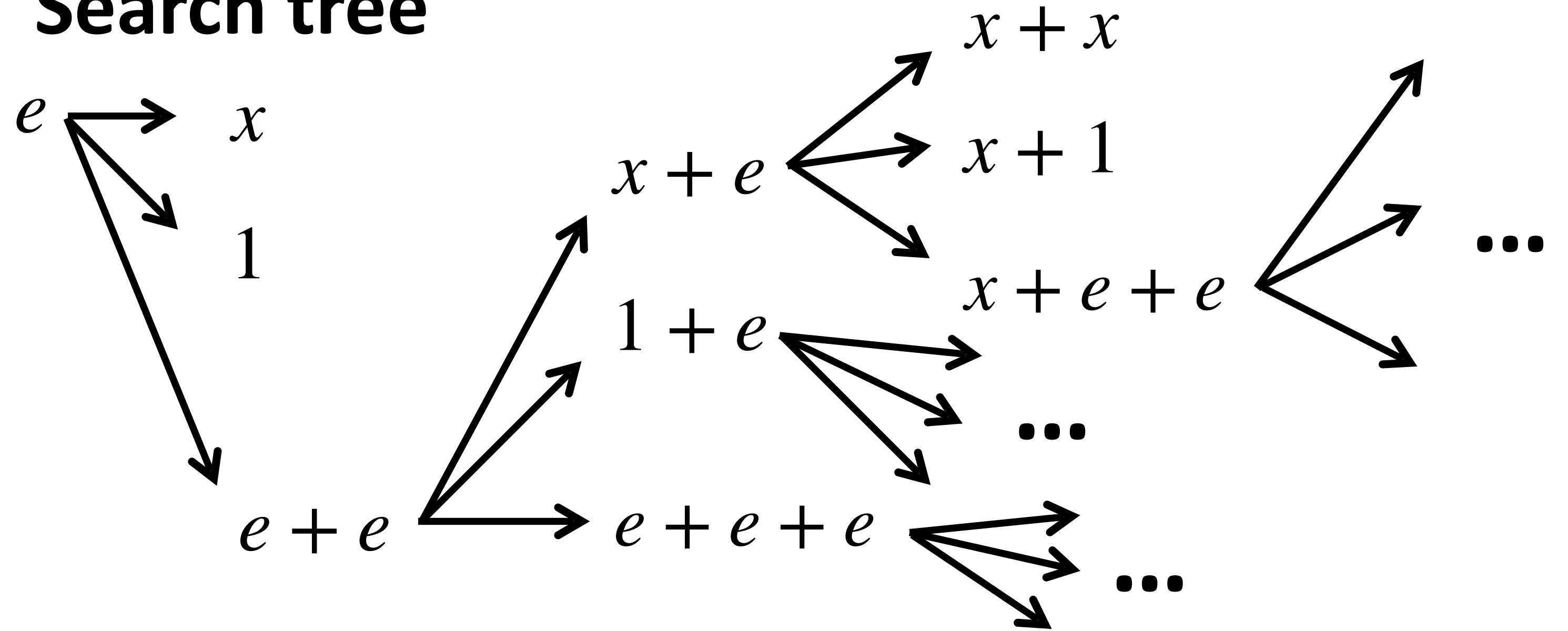
- Step 1: begin with the start symbol
- Step 2: pick a non-terminal in current result and replace it with one of its productions
- **Step 3: continue step 2 until no more non-terminal remains (i.e., only terminals)**

Top-Down Search In Action

Context-Free Grammar

$$\begin{array}{l} e \rightarrow x \\ \quad | 1 \\ \quad | e + e \end{array}$$

Search tree



How to generate this search tree?

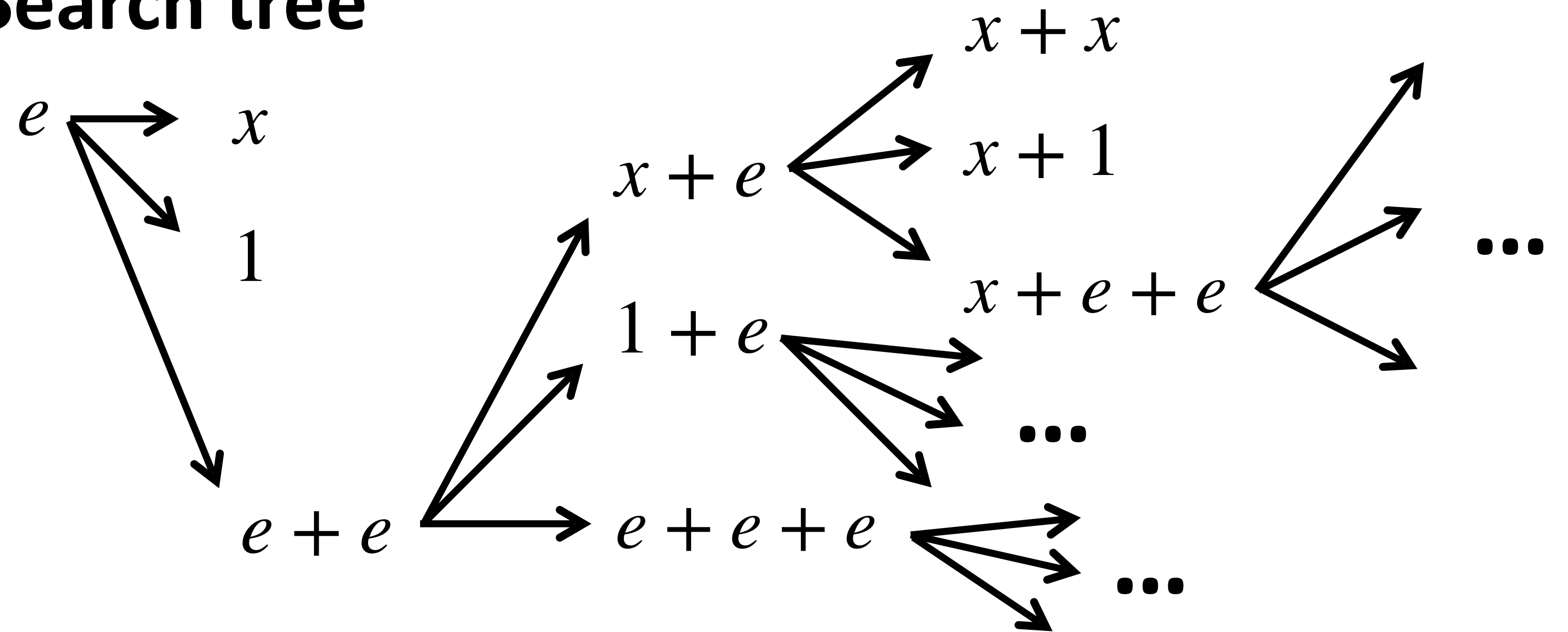
- Step 1: begin with the start symbol
- Step 2: pick a non-terminal in current result and replace it with one of its productions
- **Step 3: continue step 2 until no more non-terminal remains (i.e., only terminals)**

Top-Down Search Recap..

Context-Free Grammar

$e \rightarrow x$
| 1
| $e + e$

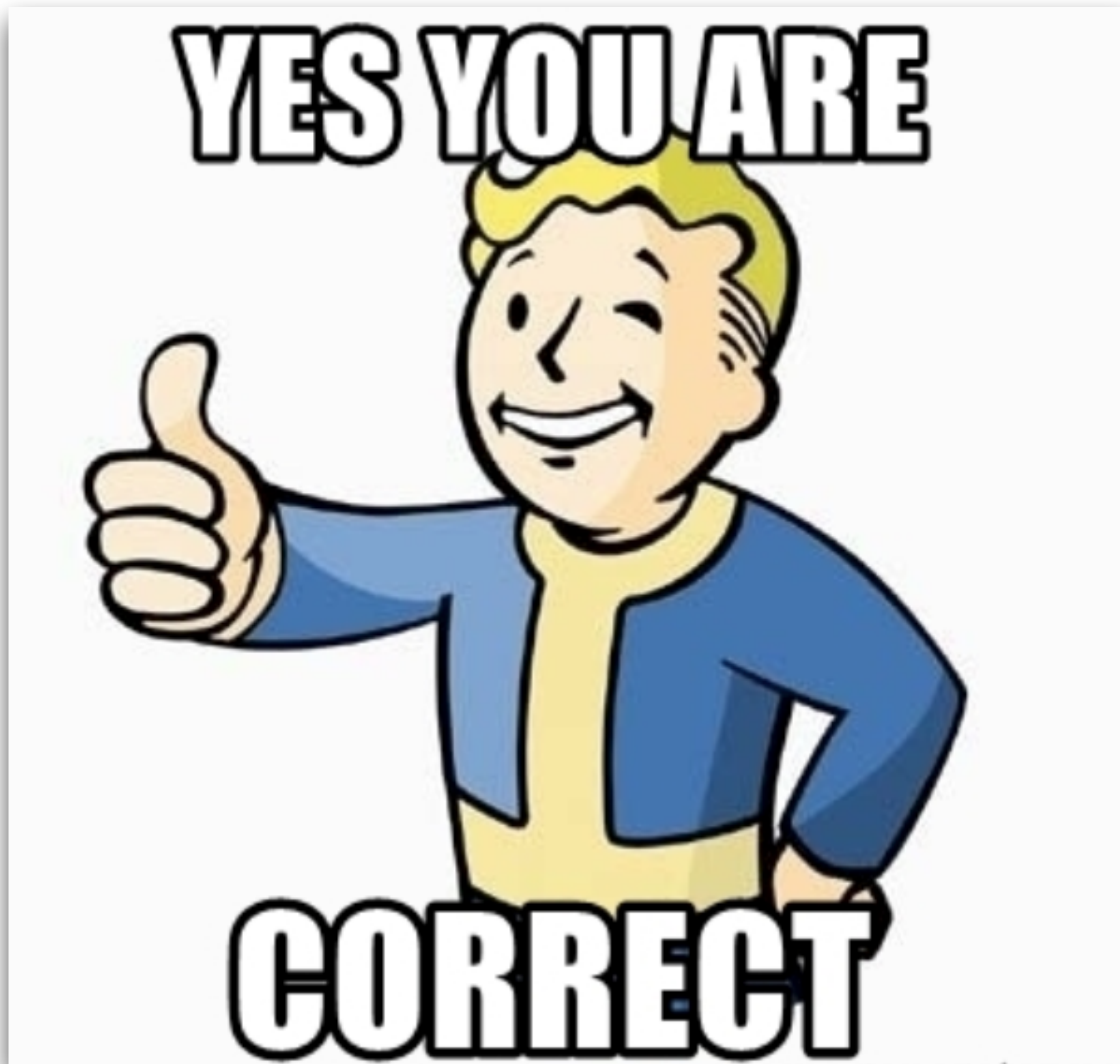
Search tree



What if my grammar gets bigger?

$e \rightarrow x$
| 1
| $e + e$
| $e - e$
| ...

Synthesis Seems Computationally Expensive?!



- Top-down search has exponential time complexity (to program size)
 - In general, program synthesis is computationally intractable, if not undecidable
- .. which means synthesis in theory does not scale to complex problems..

But, recent advances lead to empirically efficient synthesis!

State-of-the-Art

- How many lines of code can be **automatically synthesized** within a reasonable amount of time?
 - This depends on the domain, but in general, a few dozens in minutes
 - This might sound too few / too slow?

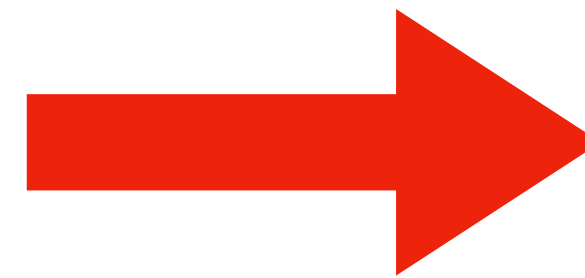
The Mythical Man-Month

- 1200 lines / year = 3 lines of code per day
 - *What?*

State-of-the-Art

To give you a concrete idea..

What WebRobot
synthesized in a few seconds



```
from traceCollector.traceCollector import TraceCollector
from selenium.common.exceptions import NoSuchElementException, StaleElementReferenceException, TimeoutException
from selenium.webdriver.common.by import By
import traceback
import os
import time

driver_path = "chromedriver"

trace_name = "wsvr1"
starting_website = "https://lv.nassaucountyny.gov/"
description = ""
"""
Given a list of tuples (section, block, lot), for each, enter them in the search, then in the search result, extract 'Address', 'Village', 'School', 'Town'
"""

# Init the trace collector and name it
trace_collector = TraceCollector(driver_path=driver_path,
                                 trace_name=trace_name,
                                 starting_website=starting_website,
                                 description=description,
                                 script_file=__file__,
                                 input_file=os.path.dirname(
                                     os.path.realpath(__file__))+"/input.json")

driver = trace_collector.get_driver()

# Sleeps for 20 seconds so that users can log in.
time.sleep(20)

# WARNING: This script requires that the user has signed up for and logged into an account (free).

for i in range(7):
    def data_path(identifier):
        return "tuples[{}].{}".format(i, identifier)

    # Sends section to appropriate box
    trace_collector.perform_send_data(
        trace_collector.get_full_xpath_from_element(
            driver.find_element_by_id("sec")
        ),
        data_path("section")
    )

    # Sends block to appropriate box
    trace_collector.perform_send_data(
        trace_collector.get_full_xpath_from_element(
            driver.find_element_by_id("blk")
        ),
        data_path("block")
    )

    # Sends lot to appropriate box
    trace_collector.perform_send_data(
        trace_collector.get_full_xpath_from_element(
            driver.find_element_by_id("lot")
        ),
        data_path("lot")
    )

    # Press "Search"
    trace_collector.perform_click(
        trace_collector.get_full_xpath_from_element(
            driver.find_element_by_id("singlebuttonsh")
        )
    )

    # Scrape address, village, school, town
    trace_collector.perform_scrape_text(
        trace_collector.get_full_xpath_from_element(
            driver.find_element_by_xpath(
                '//*[@id="right"]/section[1]/div[1]/div'
            )
        )
    )
    trace_collector.perform_scrape_text(
        trace_collector.get_full_xpath_from_element(
            driver.find_element_by_xpath(
                '//*[@id="right"]/section[1]/div[2]/div'
            )
        )
    )
    trace_collector.perform_scrape_text(
        trace_collector.get_full_xpath_from_element(
            driver.find_element_by_xpath(
                '//*[@id="right"]/section[1]/div[3]/div'
            )
        )
    )
    trace_collector.perform_scrape_text(
        trace_collector.get_full_xpath_from_element(
            driver.find_element_by_xpath(
                '//*[@id="right"]/section[1]/div[4]/div'
            )
        )
    )

    # Go back and initiate another search
    trace_collector.perform_go_back()

trace_collector.output()
```

(This is an equivalent Selenium program.)

Looking into the Future

- Synthesis as an atomic functionality in Software Engineering
 - Already come true in Excel (also “software dev”?)
 - Will come true in more and more domains
- “Programming Copilot” (ChatGPT etc.) is already here...
- Symbolic vs. Neural
 - Symbolic: search in DSL to find programs that match examples...
 - Neural: machine learning and neural nets

Looking into the Future

- Symbolic vs. Neural: both have pros and cons
 - Symbolic: highly interpretable, but slow (not scalable)..
 - Neural: generalization, but hallucinations and not interpretable..
- Many other orthogonal problems to work on
 - Usability
 - Killer apps and specialized techniques

