

Program Sketches with Scorer Holes

David Minh-Duy Cao, Sarah E. Chasins
April 16, 2025 • EPIC Retreat

The image shows a broadcast order form for FOX 4 KBTV. A callout box highlights a section of the form with the following code:

```

left = [scorer] (nodes)
right = [scorer] (nodes)
top = [scorer] (nodes)
bottom = [scorer] (nodes)

station = [scorer] (left, right, top, bottom)


```

The background form includes the following information:

- Advertiser:** REMIT TO Sinclair Broadcast Gro
- Agency:** c/o KBTV
- Address:** PO Box 206270, Dallas, TX 75320-6270
- Product:** Special Handling
- Brand:** Product
- Station:** FOX 4 KBTV
- Contract Line Remarks:** 2a news
- Schedule:** 02/26/25 7:31AM (We)

Motivation:

Folks are often charged with getting information out of **templated** PDF documents



REMIT TO Sinclair Broadcast Gro
c/o KBTV

station_name

Advertiser Pete Buttigieg for President-D (132717)
Agency AKPD Message & Media (6417)
Buyer PARETS,LYNDA
Salesperson Millennium/CHI, Chicago (1094)
 ph: (312) 832-6600

Special Handling None
Product POLITICAL CANDIDATE (ns) (1186)
Brand PETE FOR AMERICA (1249532)
Acct Types National/Political Candidate Agency BRD
Est/External # /ECR26808399
Demo A35+
Revision
Comments Pete for America laura heinrich
 only booking Wed 2/26 for right now. Will

Invoice 7294160
Inv Date 3/1/2020
Terms CIA
Contract 4257797
Bill Type Weekly/irregular
Period 2/24/2020 - 3/1/2020

AKPD Message & Media
 730 N Franklin
 Ste 404
 Chicago, IL 60654

AgM


Beaumont (KBTV)

OFFICIAL BILLING INVOICE

Line	Type	Scheduled	Schedule Days to Run	Air Time	Length	Copy/ISCI	Amount	Remarks
1.0	Contract Line Remark:		7a news					
	News	220211-News-The Morn	Day,SP-1	02/26/20 7:31AM (We)	00:30	PFA 20 036H	\$55.00	
2.0	Contract Line Remark:		8a news					
	News	220212-News-The Morn	Day,SP-1	02/26/20 8:21AM (We)	00:30	PFA 20 036H	\$55.00	

Motivation:

Folks are often charged with getting information out of **templated** PDF documents that are **heterogeneous** (differences in formatting or alignment, PDF artifacts, ...).



REMIT TO Sinclair Broadcast Gro
 c/o KBTV
 P.O. Box 200270
 Dallas, TX 75320-6270

Advertiser Pete Buttigieg for President-D (132717)
Agency AKPD Message & Media (6417)
Buyer PARETS,LYNDA
Salesperson Millennium/CHI, Chicago (1094)
 ph: (312) 832-6600

Special Handling None
Product POLITICAL CANDIDATE (ns) (1186)
Brand PETE FOR AMERICA (1249532)
Acct Types National/Political Candidate Agency BRD
Est/External # /ECR26808399
Demo A35+
Revision
Comments Pete for America laura heinrich
 only booking Wed 2/26 for right now. Will

Invoice 7294160
Inv Date 3/1/2020
Terms CIA
Contract 4257797
Bill Type Weekly/Irregular
Period 2/24/2020 - 3/1/2020

CO-OP/Order Type No/Normal
Package
Gen. Date 3/2/2020 1:41:49PM

AKPD Message & Media
 730 N Franklin
 Ste 404
 Chicago, IL 60654

AgM

Beaumont (KBTV) **OFFICIAL BILLING INVOICE**

Line	Type	Scheduled	Schedule Days to Run	Air Time	Length	Copy/ISCI	Amount	Remarks
1.0	Contract Line Remarks:		7a news					
	News	329211-News-The Morn	Day/1	02/26/20 7:31AM (We)	00:30	PFA 20 636H	\$55.00	
2.0	Contract Line Remarks:		8a news					
	News	329212-News-The Morn	Day/1	02/26/20 8:21AM (We)	00:30	PFA 20 636H	\$55.00	



WPXI
 P.O. Box 640132
 Pittsburgh, PA 15264-0132
 ph: (412) 237-1100
 fx: (412) 237-2247

Advertiser Pol/T Murphy/R/Con/PA (16299)
Agency BrabenderCox (3844)
Buyer BRABENDER,LIZ
Salesperson Telerep/Washington DC, Washington DC (1040)
 ph: (555) 555-5555
Product Political - Federal Candidate (1067)
Brand POL/T MURPHY/R/CON/PA (29258)
Acct Types National/Political
Est/Headline /06292001
Demo
Revision
Comments MURPHY 10/23
 ***** THIS IS A CASH IN ADVANCE

Invoice 640565
Inv Date 10/28/2012
Terms Net 30
Contract 331377
Bill Type Standard
Period 10/1/2012 - 10/28/2012

CO-OP/Order Type Yes/Normal
Package
Gen. Date 10/29/2012 1:23:10PM

BrabenderCox
 1218 Grandview Avenue
 1st Floor
 Pittsburgh, PE 15211

AgM

Pittsburgh (WPXI) **OFFICIAL BILLING INVOICE**

Line	Type	Scheduled	Schedule Days to Run	Air Time	Length	Program	Copy/ISCI	Amount	Remarks
22.0	Contract Line Remarks:		TODAY 7-9A/112						
	Spot	7:00:00AM- 8:00:00	Day,W-L,Th-L,F-1	10/24/12 7:24AM (We)	00:30	Today Show	BCTM1203H	\$625.00	
	Spot	7:00:00AM- 8:00:00	Day,W-L,Th-L,F-1	10/26/12 7:27AM (Fr)	00:30	Today Show	BCTM1203H	\$625.00	
22.0	Spot			10/25/12			i preempt @	\$0.00	Advertiser Instructions
24.0	Contract Line Remarks:		JUDY 4-430P/114						

Motivation:

Folks are often charged with getting information out of **templated** PDF documents that are **heterogeneous** (differences in formatting or alignment, PDF artifacts, ...) and **numerous** (hundreds or thousands of documents).

FOX 4 KBTV

REMIT TO Sindair Broadcast Gro
c/o KBTV
PO Box 999970
Dallas, TX 75320-6270

Advertiser: Pete Buttigieg for President-D (132717)
Agency: AKPD Message & Media (6417)
Buyer: PARETS,LYNDA
Salesperson: Millennium/CHI, Chicago (1094)
Special Handling: None
Product: POLITICAL CANDIDATE (ns) (1186)
Brand: PETE FOR AMERICA (1249532)
Acct Types: National/Political Candidate Agency BRD
Est/External #: /ECR26808399
Demo: A35+
Revision:
Comments: Pete for America laura heinrich
only booking Wed 2/26 for right now. Will

Invoice: 7294160
Inv Date: 3/1/2020
Terms: CIA
Contract: 4257797
Bill Type: Weekly/Irregular
Period: 2/24/2020 - 3/1/2020

AKPD Message & Media
730 N Franklin
Ste 404
Chicago, IL 60654

Beaumont (KBTV)

OFFICIAL BILLING INVOICE

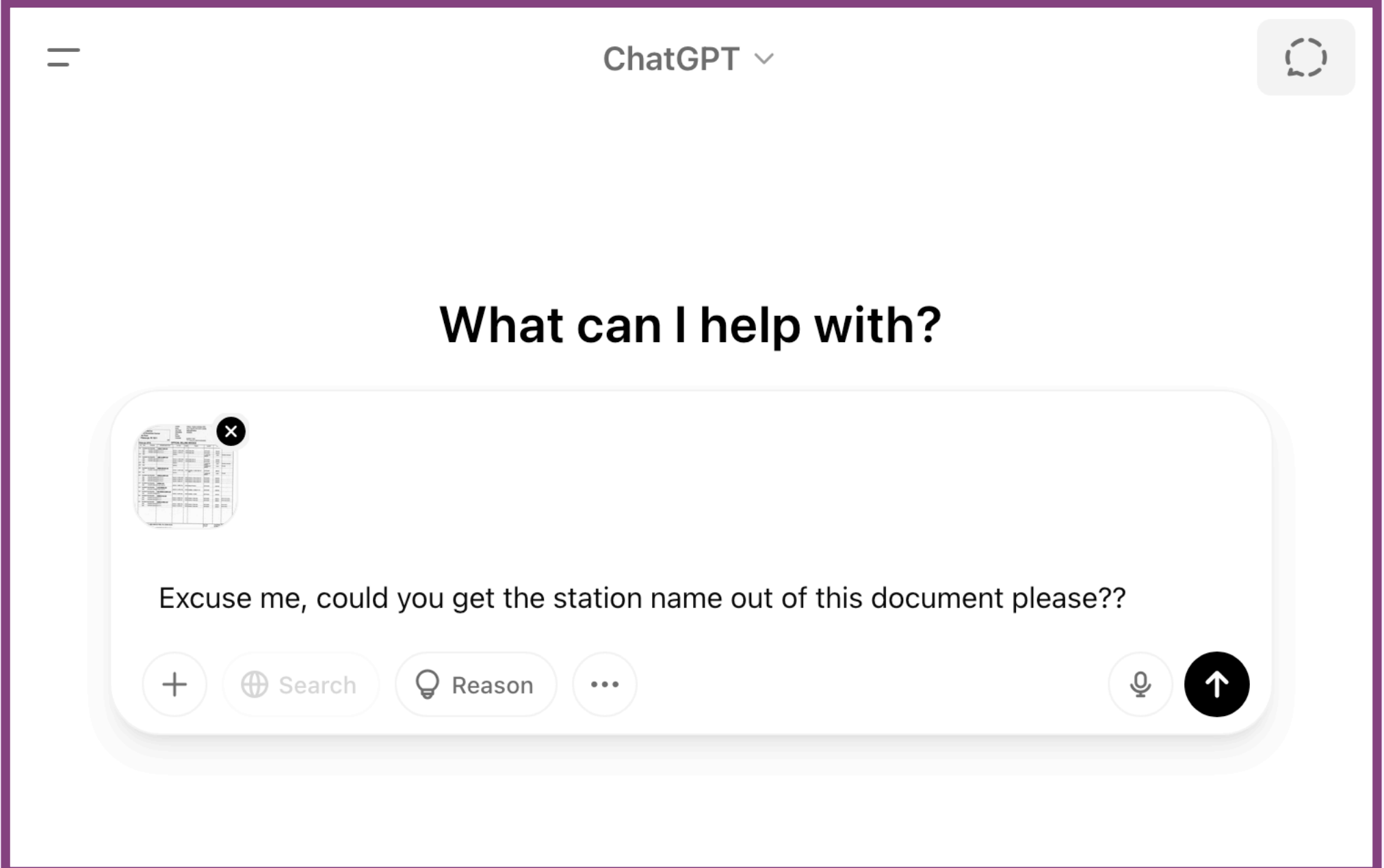
Line	Type	Scheduled	Schedule Days to Run	Air Time	Length	Copy/ISCI	Amount	Remarks
1.0	Contract Line Remark:		<u>7a news</u>					
	News	3/2/20	3/2/20	7:31AM (W)	00:30	RFA 20 036H	\$55.00	
2.0	Contract Line Remark:		<u>8a news</u>					
	News	3/2/20	3/2/20	8:21AM (W)	00:30	RFA 20 036H	\$55.00	
22.0	Spot			10/25/12				
24.0	Contract Line Remark:		<u>JUDY 4-430P/114</u>			1 preempt @ \$625.00	\$0.00	Advertiser Instructions

and 1,225 more

Under-resourced folks lack the tools to process messy documents at scale.

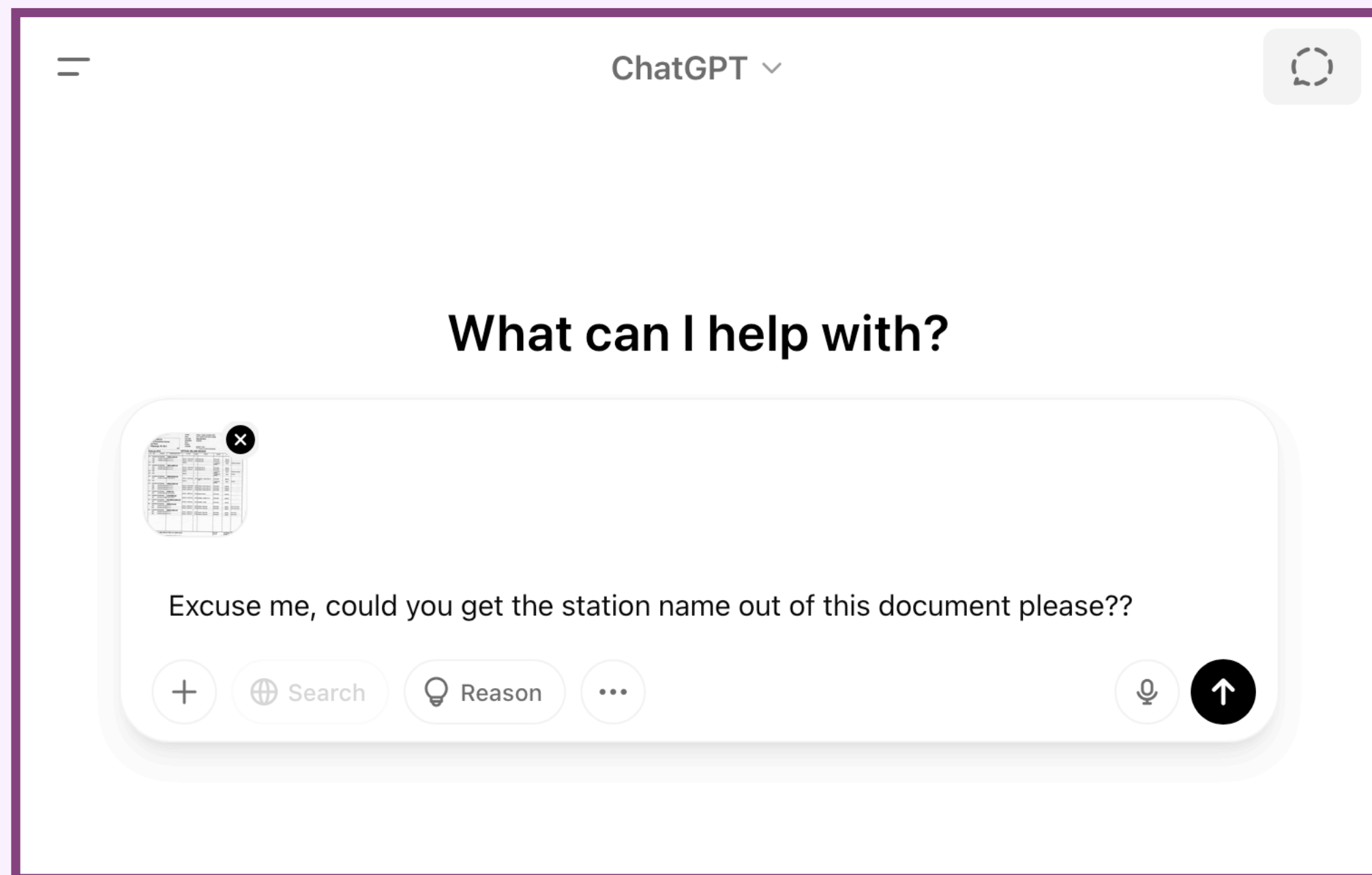
Under-resourced folks lack the tools to process messy documents at scale.

LLMs (e.g., GPT-4o)?



Under-resourced folks lack the tools to process messy documents at scale.

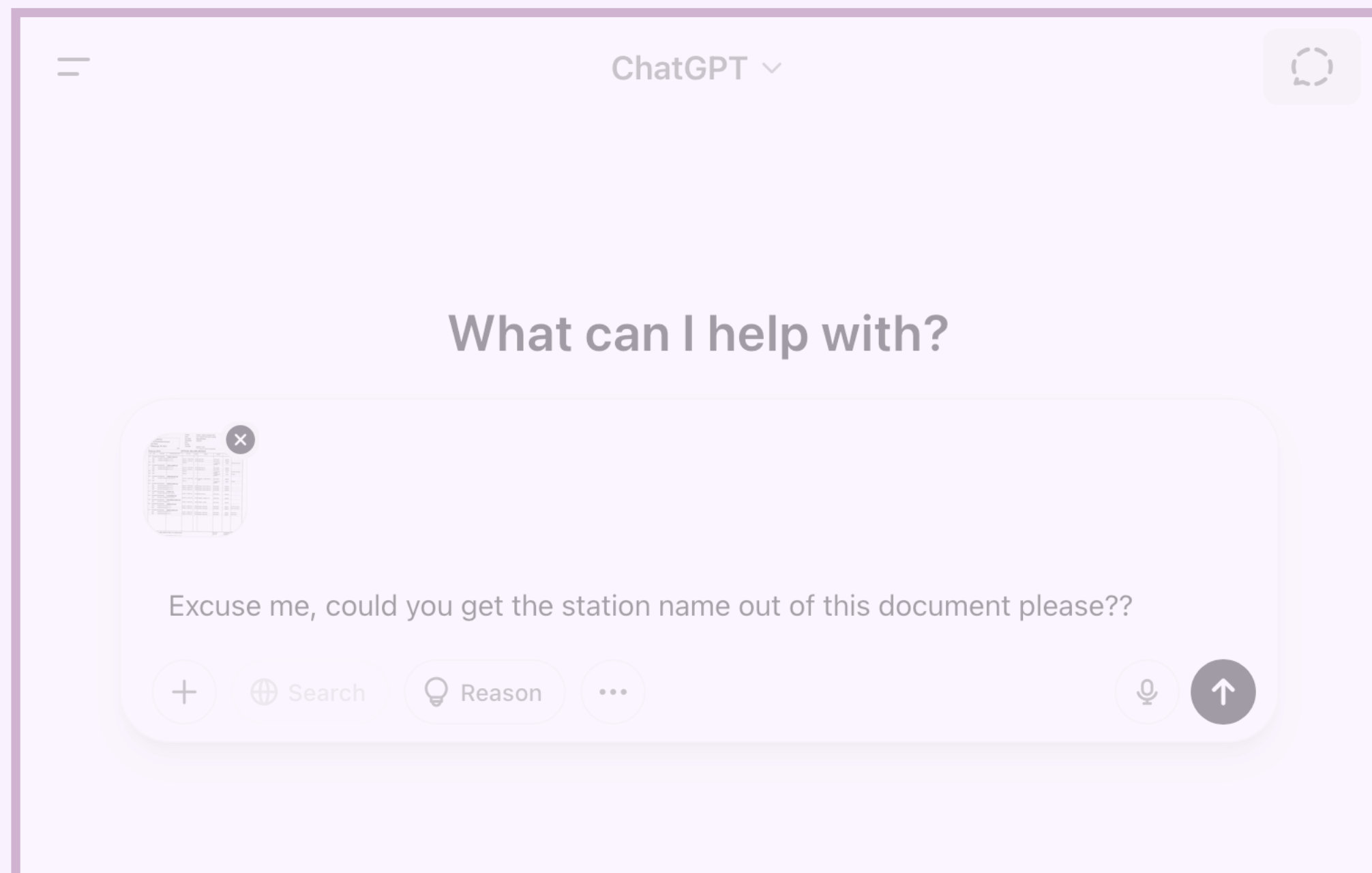
LLMs (e.g., GPT-4o)?



Running our benchmark suite of **1,227 documents** took **≥14.5 hours** and cost us **\$1,675.05**.

Under-resourced folks lack the tools to process messy documents at scale.

LLMs (e.g., GPT-4o)?



Traditional program synthesis?

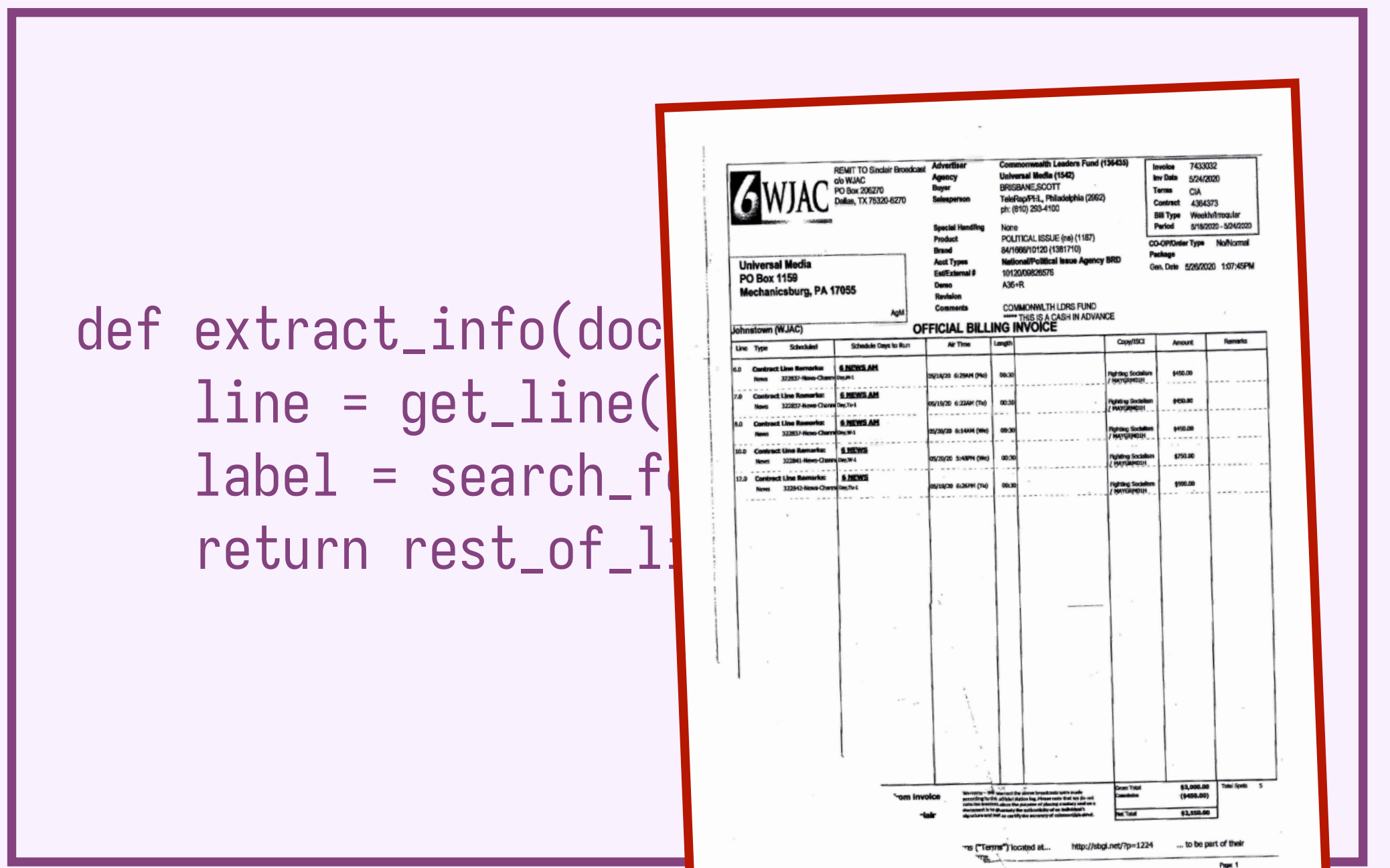
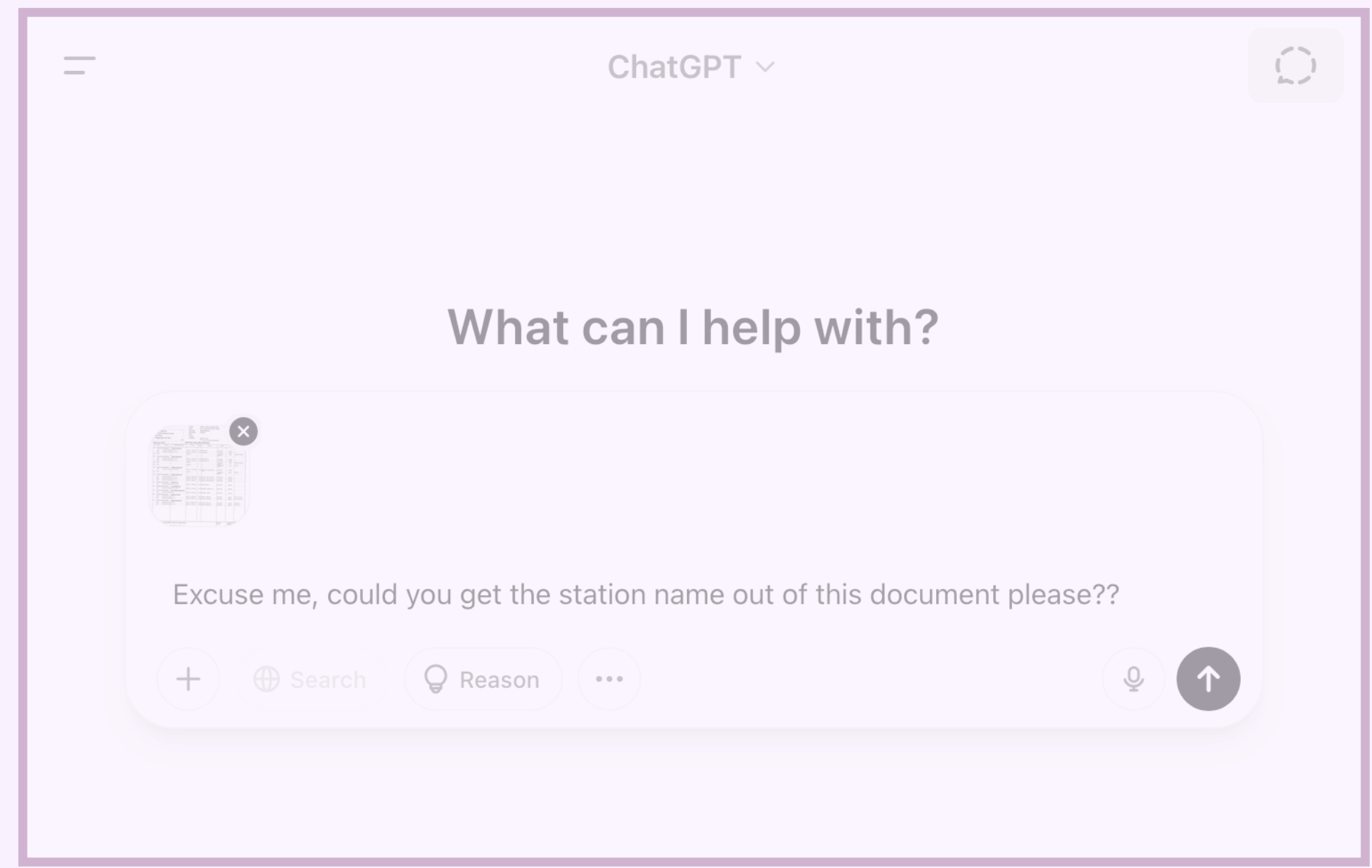
```
def extract_info(doc):  
    line = get_line(4, doc)  
    label = search_for("Station:", line)  
    return rest_of_line(label)
```

Running our benchmark suite of **1,227 documents** took **≥14.5 hours** and cost us **\$1,675.05**.

Under-resourced folks lack the tools to process messy documents at scale.

LLMs (e.g., GPT-4o)?

Traditional program synthesis?




Running our benchmark suite of **1,227 documents** took \geq **14.5 hours** and cost us **\$1,675.05**.

Traditional synthesis techniques **don't work** with messy PDF data.

Can we adapt traditional synthesis techniques for messy data domains?

Can we adapt traditional synthesis techniques for messy data domains?
Let's use insights from Ringer and web automation.

Electronics > Camera & Photo > Digital Cameras > Mirrorless Cameras



Roll over image to zoom in

Nikon Z6 III | Full-Frame mirrorless Stills/Video Camera with 6K/60p Internal RAW Recording | Nikon USA Model

[Visit the Nikon Store](#)
4.4 ★★★★★ (59) | [Search this page](#)
100+ bought in past month


\$2,496⁹⁵

Or **\$98.82**/mo (48 mo). Select from 1 plan

FREE Returns ▾
Save up to 2% with business pricing. Sign up for a free Amazon Business account

Pay **\$208.08/month for 12 months, interest-free upon approval for Amazon Visa**

Electronics > Camera & Photo > Digital Cameras > Mirrorless Cameras



Roll over image to zoom in

Canon EOS R6 Mark II Mirrorless Camera (Body Only), Full-Frame Camera, 24.2 Megapixel CMOS Sensor, Photo and Video Capabilities, Black

[Visit the Canon Store](#)
4.8 ★★★★★ (544) | [Search this page](#)
Amazon's Choice Overall Pick
500+ bought in past month

-20% \$1,999⁰⁰

List Price: ~~\$2,499.00~~

Or **\$79.11**/mo (48 mo). Select from 1 plan


FREE Returns ▾

Can we adapt traditional synthesis techniques for messy data domains?

Let's use insights from Ringer and web automation.

Ringer's **scorers** contend with webpage heterogeneity.

Electronics > Camera & Photo > Digital Cameras > Mirrorless Cameras



Roll over image to zoom in

Nikon Z6 III | Full-Frame mirrorless Stills/Video Camera with 6K/60p Internal RAW Recording | Nikon USA Model

[Visit the Nikon Store](#)

4.4 ★★★★★ (59) | [Search this page](#)

100+ bought in past month

\$2,496⁹⁵


Or **\$98.82**/mo (48 mo). Select from 1 plan

FREE Returns ▼

Save up to 2% with business pricing. Sign up for a free Amazon Business account

Pay **\$208.08/month for 12 months, interest-free upon approval for Amazon Visa**

Electronics > Camera & Photo > Digital Cameras > Mirrorless Cameras



Roll over image to zoom in

Canon EOS R6 Mark II Mirrorless Camera (Body Only), Full-Frame Camera, 24.2 Megapixel CMOS Sensor, Photo and Video Capabilities, Black

[Visit the Canon Store](#)

4.8 ★★★★★ (544) | [Search this page](#)

Amazon's Choice Overall Pick

500+ bought in past month

-20% \$1,999⁰⁰

List Price: ~~\$2,499.00~~


Or **\$79.11**/mo (48 mo). Select from 1 plan

FREE Returns ▼

Ringer's scorers contend with webpage heterogeneity.

Unseen webpage

Electronics › Camera & Photo › Digital Cameras › Mirrorless Cameras



Nikon Z6 III | Full-Frame mirrorless Stills/Video Camera with 6K/60p Internal RAW Recording | Nikon USA Model

[Visit the Nikon Store](#)

4.4 ★★★★★ (59) | [Search this page](#)

100+ bought in past month

\$2,496⁹⁵

Or **\$98.82** /mo (48 mo). [Select from 1 plan](#)

FREE Returns ▾

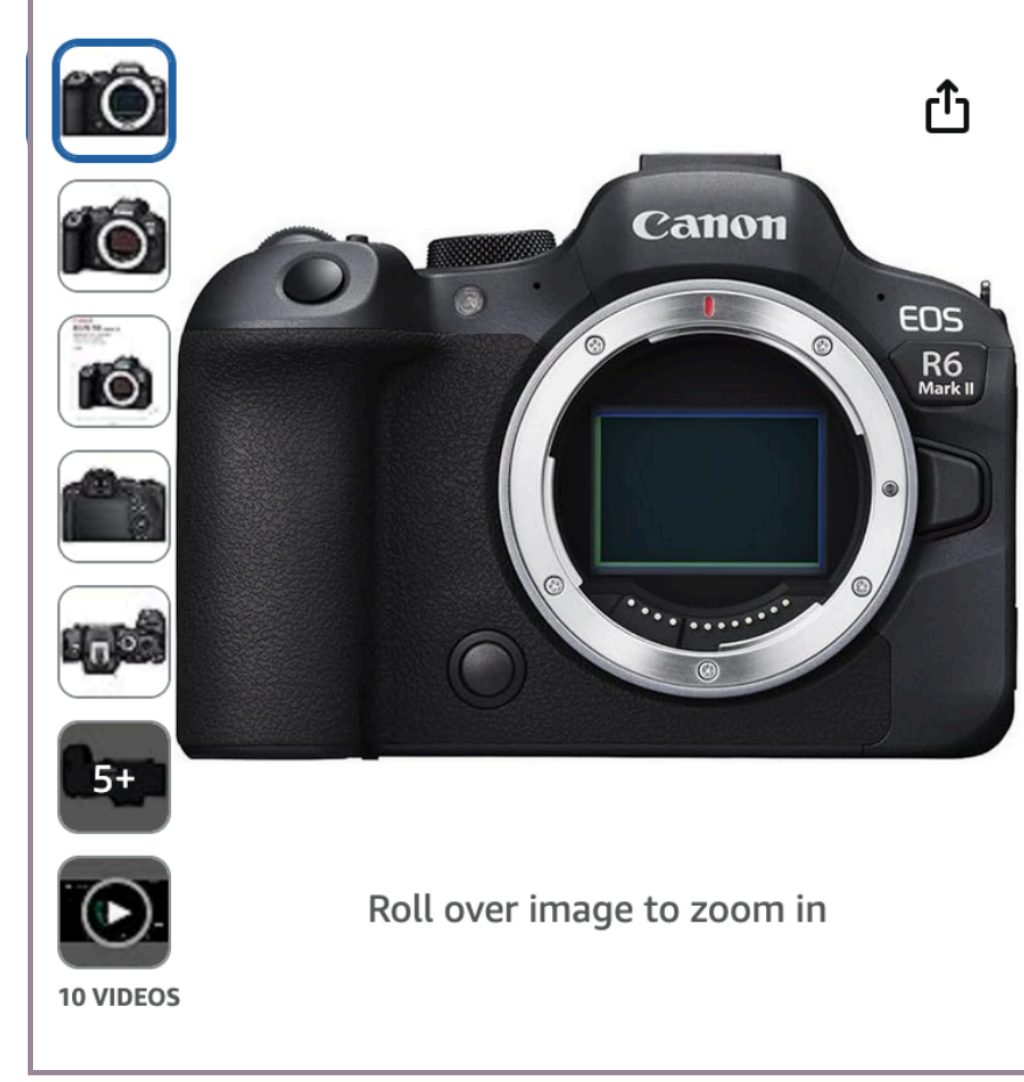
Save up to 2% with business pricing. Sign up for a free Amazon Business account

Pay \$208.08/month for 12 months, interest-free upon approval for Amazon Visa

6+
6 VIDEOS

Roll over image to zoom in

Electronics › Camera & Photo › Digital Cameras › Mirrorless Cameras



Canon EOS R6 Mark II Mirrorless Camera (Body Only), Full-Frame Camera, 24.2 Megapixel CMOS Sensor, Photo and Video Capabilities, Black

[Visit the Canon Store](#)

4.8 ★★★★★ (544) | [Search this page](#)

Amazon's Choice Overall Pick

500+ bought in past month

-20% \$1,999⁰⁰

List Price: ~~\$2,499.00~~

Or **\$79.11** /mo (48 mo). [Select from 1 plan](#)

FREE Returns ▾

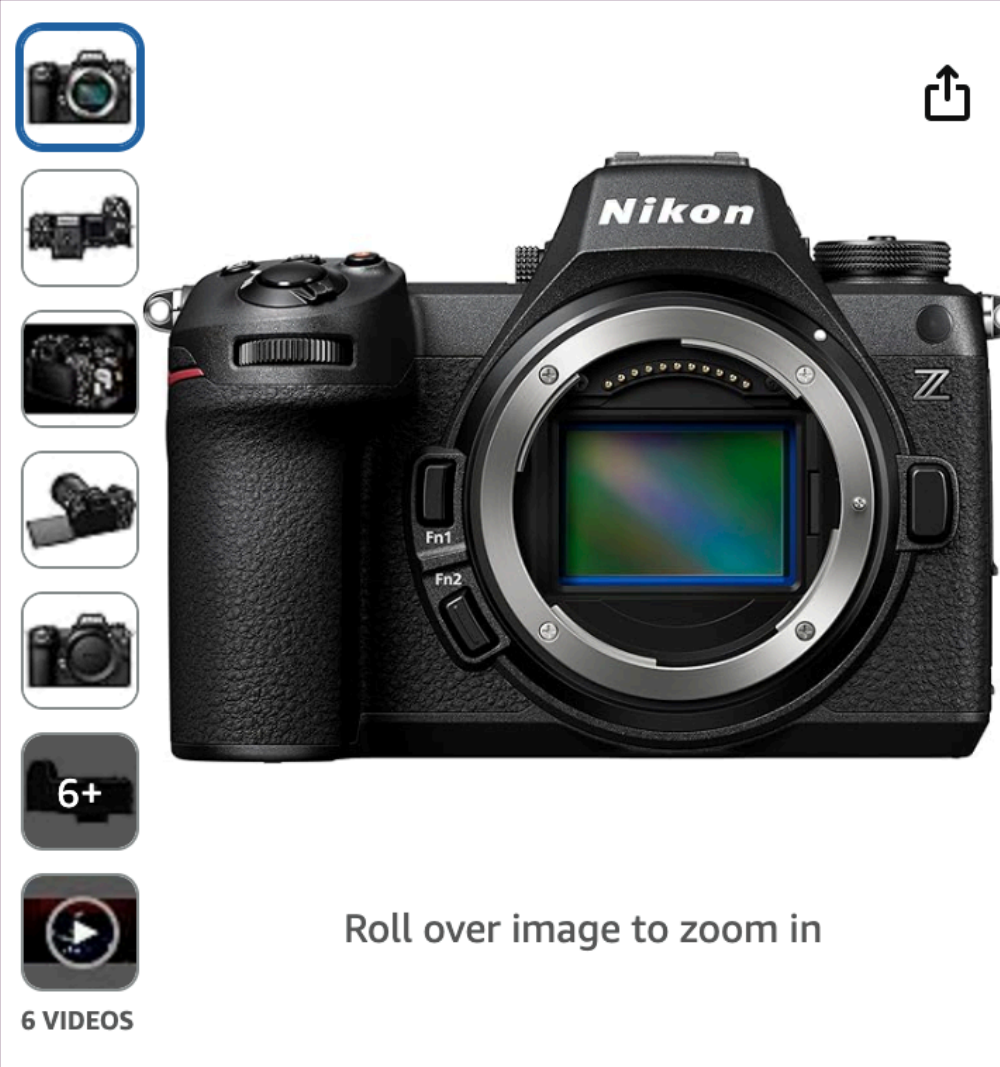
5+
10 VIDEOS

Roll over image to zoom in

User-provided example

Ringer's scorers contend with webpage heterogeneity.

Electronics > Camera & Photo > Digital Cameras > Mirrorless Cameras



Nikon Z6 III | Full-Frame mirrorless Stills/Video Camera with 6K/60p Internal RAW Recording | Nikon USA Model

[Visit the Nikon Store](#)

4.4 ★★★★★ (59) | [Search this page](#)

100+ bought in past month

\$2,496⁹⁵

Or **\$98.82** /mo (48 mo). [Select from 1 plan](#)

FREE Returns ▾

Save up to 2% with business pricing. Sign up for a free Amazon Business account

Pay **\$208.08/month for 12 months, interest-free upon approval for Amazon Visa**

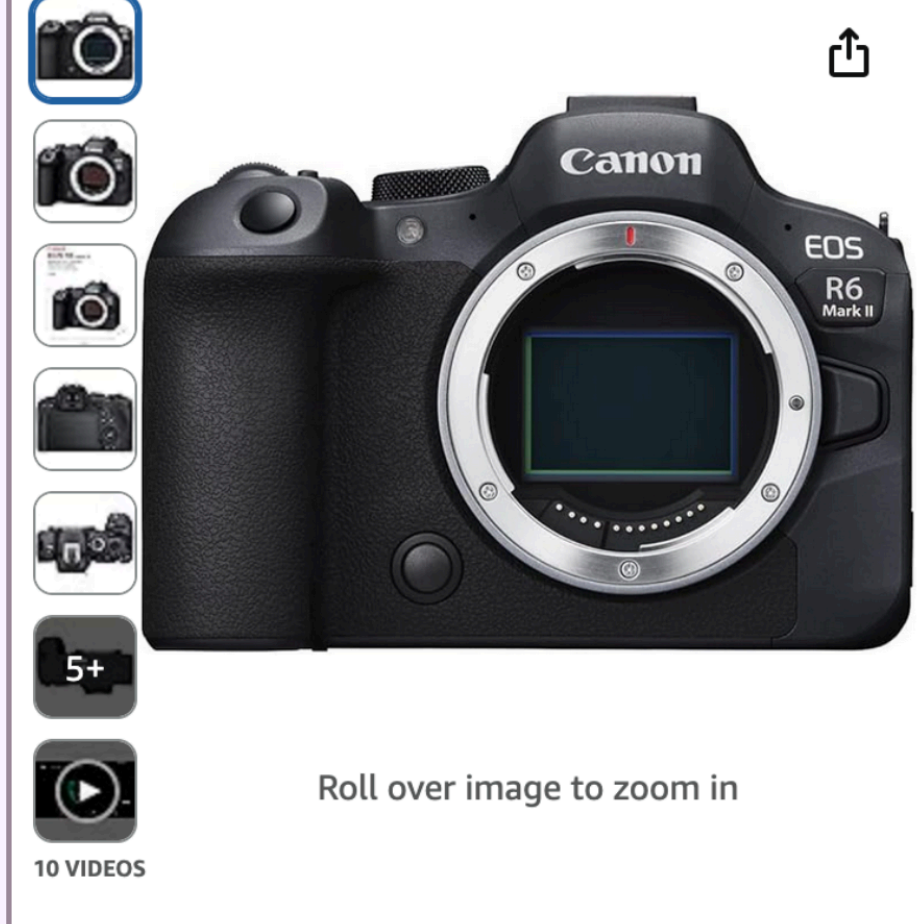
Roll over image to zoom in

6 VIDEOS

User-provided example

Electronics > Camera & Photo > Digital Cameras > Mirrorless Cameras

Canon EOS R6 Mark II Mirrorless Camera (Body Only), Full-Frame Camera, 24.2 Megapixel CMOS Sensor, Photo and Video Capabilities, Black



[Visit the Canon Store](#)

4.8 ★★★★★ (544)

[Search this page](#)

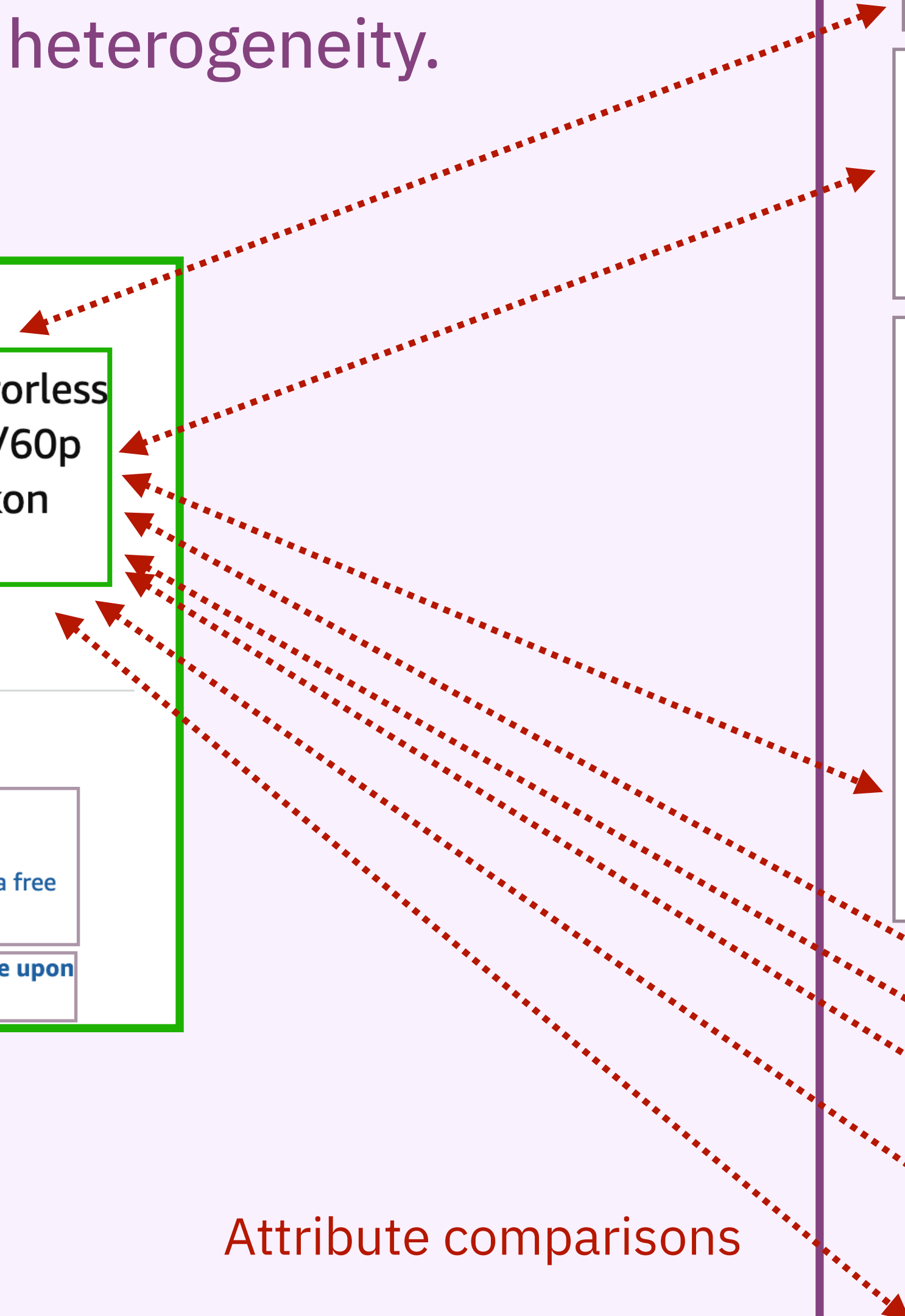
-20% \$1,999⁰⁰
List Price: \$2,499.00 ⓘ

Or **\$79.11** /mo (48 mo). [Select from 1 plan](#)

FREE Returns ▾

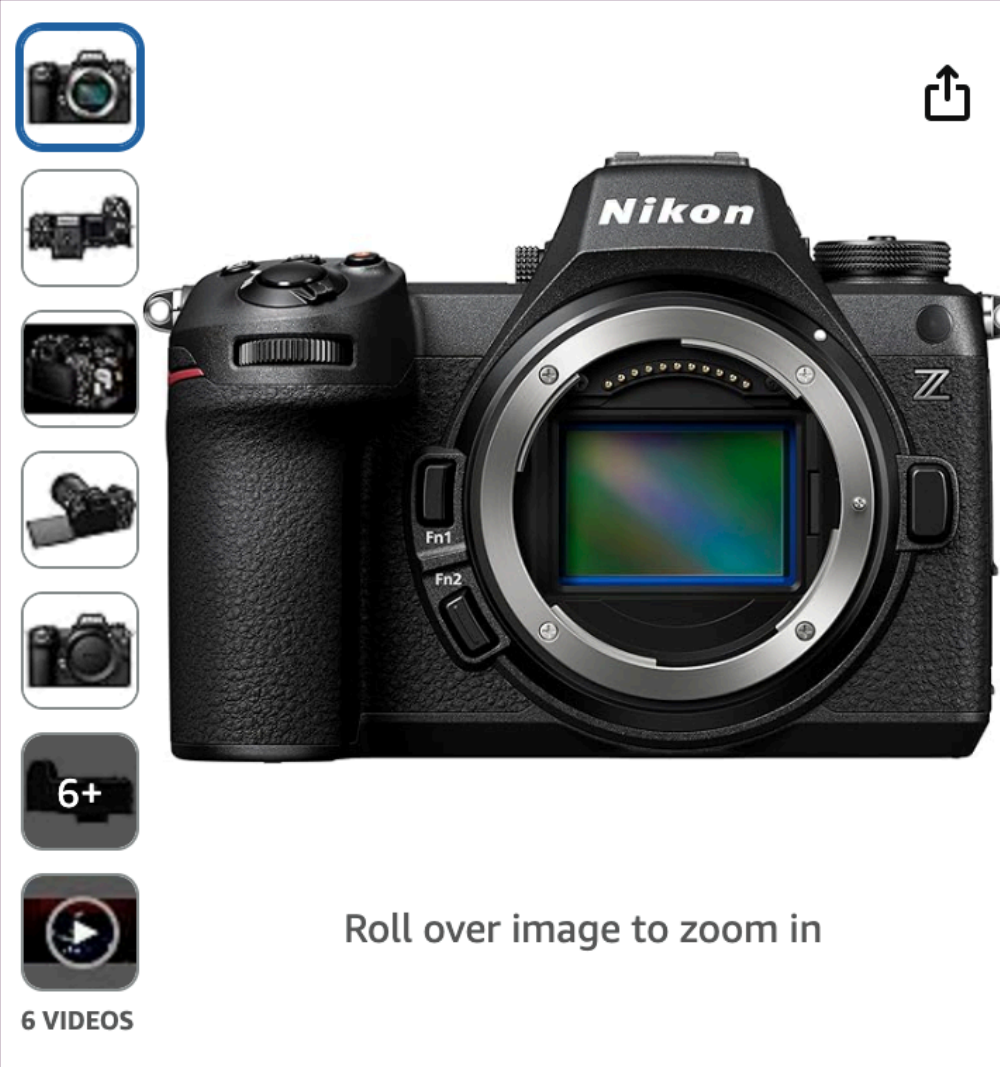
⋮

Attribute comparisons



Ringer's scorers contend with webpage heterogeneity.

Electronics > Camera & Photo > Digital Cameras > Mirrorless Cameras



Nikon Z6 III | Full-Frame mirrorless Stills/Video Camera with 6K/60p Internal RAW Recording | Nikon USA Model

Visit the Nikon Store

4.4 ★★★★★ (59) | Search this page

100+ bought in past month

\$2,496⁹⁵

Or **\$98.82** /mo (48 mo). Select from 1 plan

FREE Returns ▾

Save up to 2% with business pricing. Sign up for a free Amazon Business account

Pay **\$208.08/month for 12 months, interest-free upon approval for Amazon Visa**

User-provided example


Same tag? ✗
 Same ID? ✓
 Same classes? ✓
 Common attrs: 2

Attribute comparisons

Unseen webpage

Electronics > Camera & Photo > Digital Cameras > Mirrorless Cameras

Canon EOS R6 Mark II Mirrorless Camera (Body Only), Full-Frame Camera, 24.2 Megapixel CMOS Sensor, Photo and Video Capabilities, Black



Visit the Canon Store

4.8 ★★★★★ (544)

Search this page

-20% \$1,999⁰⁰
List Price: \$2,499.00 ⓘ

Or **\$79.11** /mo (48 mo). Select from 1 plan

FREE Returns ▾

⋮

Ringer's scorers contend with webpage heterogeneity.
But they require structured documents.

Hundreds or
thousands of
nodes

Unseen webpage

Electronics › Camera & Photo › Digital Cameras › Mirrorless Cameras

Canon EOS R6 Mark II Mirrorless Camera (Body Only), Full-Frame Camera, 24.2 Megapixel CMOS Sensor, Photo and Video Capabilities, Black



Visit the Canon Store

4.8 ★★★★★ (544)

Search this page

-20% \$1,999⁰⁰

List Price: \$2,499.00 ⓘ

Or \$79.11 /mo (48 mo). Select from 1 plan

FREE Returns ▾



Ringer's scorers require structured documents.



REMIT TO Sinclair Broadcast
45 WXLV
PO Box 206270
Dallas, TX 75320-6270

Advertiser Citizens for Responsible Energy (84874)
Agency Matsori Media LLC (19039)
Buyer MediaAssistant
Salesperson Millennium(DC, Washington, DC) (1108)
ph: (202) 955-5342 | fx: (202) 955-5348x

Special Handling None
Product POLITICAL ISSUE (ns) (1187)
Brand NCHD29 (1340882)
Acct Types National Political Issue Agency BRD
Est/External # 11637ECR26803967
Demo
Revision
Comments WXLV
NCHD29

Invoice 7245066
Inv Date 2/23/2020
Terms CIA
Contract 4248847
Bill Type Weekly/Regular
Period 2/23/2020 - 3/23/2020

CO-OP/Order Type Nq/Normal
Package
Gen Date 2/25/2020 8:23:41AM

Matsori Media LLC
1201 Hampton Street
Ste 3B
Columbia, SC 29201

Greensboro (WXLV) OFFICIAL BILLING INVOICE

Line	Type	Scheduled	Schedule (Days to Run)	Air Time	Length	Program	Copy/ESCI	Amount	Remarks
1.0	Contract Line								
1.0	News	8:30P	ABC America	02/19/20	5:15AM (Thu)	00:30	ABC America This Morning	SEP TV021820H	\$20.00
1.0	News	8:30P	ABC America	02/19/20	5:58AM (Thu)	00:30	ABC America This Morning	SEP TV021820H	\$20.00
1.0	News	8:30P	ABC America	02/20/20	5:23AM (Fri)	00:30	ABC America This Morning	SEP TV021820H	\$20.00
1.0	News	8:30P	ABC America	02/20/20	5:59AM (Fri)	00:30	ABC America This Morning	SEP TV021820H	\$20.00
1.0	News	8:30P	ABC America	02/21/20	5:10AM (Sat)	00:30	ABC America This Morning	SEP TV021820H	\$20.00
1.0	News	8:30P	ABC America	02/21/20	5:54AM (Sat)	00:30	ABC America This Morning	SEP TV021820H	\$20.00
4.0	SPOT	10:14	ABC Good Mon	02/19/20	7:59AM (Thu)	00:30	ABC Good Morning America	SEP TV021820H	\$90.00
4.0	SPOT	10:14	ABC Good Mon	02/19/20	8:59AM (Thu)	00:30	ABC Good Morning America	SEP TV021820H	\$90.00
4.0	SPOT	10:14	ABC Good Mon	02/20/20	7:59AM (Fri)	00:30	ABC Good Morning America	SEP TV021820H	\$90.00
4.0	SPOT	10:14	ABC Good Mon	02/20/20	8:59AM (Fri)	00:30	ABC Good Morning America	SEP TV021820H	\$90.00
4.0	SPOT	10:14	ABC Good Mon	02/21/20	7:59AM (Sat)	00:30	ABC Good Morning America	SEP TV021820H	\$90.00
4.0	SPOT	10:14	ABC Good Mon	02/21/20	8:59AM (Sat)	00:30	ABC Good Morning America	SEP TV021820H	\$90.00
7.0	SPOT	10:14	ABC Good Mon	02/22/20	8:29AM (Sun)	00:30	ABC Good Morning America	SEP TV021820H	\$30.00
8.0	SPOT	10:14	ABC Good Mon	02/22/20	9:29AM (Sun)	00:30	ABC Good Morning America	SEP TV021820H	\$30.00
11.0	SPOT	14:34	ABC Good Mon	02/23/20	8:29AM (Sun)	00:30	ABC Good Morning America	SEP TV021820H	\$40.00

Net 30 Payment is due within 30 days from invoice date.
Station owned or provided services by the Sinclair Broadcast Group.

Gross Total \$760.00
Commission \$114.00
Net Total \$646.00

Total Spots 13

Officially printed on 02/27/20 11:44AM Page 8

Hundreds or thousands of nodes

Unseen webpage

Electronics > Camera & Photo > Digital Cameras > Mirrorless Cameras

Canon EOS R6 Mark II Mirrorless Camera (Body Only), Full-Frame Camera, 24.2 Megapixel CMOS Sensor, Photo and Video Capabilities, Black



Roll over image to zoom in

10 VIDEOS

Visit the Canon Store

4.8 ★★★★★ (544)

Search this page

-20% \$1,999⁰⁰


List Price: \$2,499.00

Or \$79.11 /mo (48 mo). Select from 1 plan

FREE Returns



Ringer's scorers require structured documents.



REMIT TO Sinclair Broadcast
45 WXLV
PO Box 206270
Dallas, TX 75320-6270

Advertiser Citizens for Responsible Energy (84874)
Agency Matsori Media LLC (19039)
Buyer MediaAssistant1
Salesperson Millennium DC, Washington DC (1108)
ph: (202) 955-5342 | fx: (202) 955-5348x

Invoice 7245066
Inv Date 2/23/2020
Terms C/A
Contract 4248847
Bill Type Weekly Regular
Period 2/17/2020 - 2/23/2020

Matsori Media LLC
1201 Hampton Street
Ste 3B
Columbia, SC 29201

Special Handling None
Product POLITICAL ISSUE (1187)
Brand NCHD29 (1340882)
Acct Types National Political Issue Agency BRD
Est/External # 11637/ECR26803967
Demo

CO-OP/Order Type Normal
Package
Gen Date 2/25/2020 2:23:41 AM

Revision WXLV
Comments WXLV

OFFICIAL BILLING INVOICE

Line	Type	Scheduled	Schedule (Days to Run)	Air To	Length	Program	Copy/ESCI	Amount	Remarks
1.0	Contract Line	Contract Line Remarks							
1.0	News	8:30P	ABC America	---M--	02/19/20	5:30P	ABC America This Morning	SEP TV021820H	\$20.00
1.0	News	8:30P	ABC America	---M--	02/19/20	5:30P	ABC America This Morning	SEP TV021820H	\$20.00
1.0	News	8:30P	ABC America	---M--	02/20/20	5:30P	ABC America This Morning	SEP TV021820H	\$20.00
1.0	News	8:30P	ABC America	---M--	02/20/20	5:30P	ABC America This Morning	SEP TV021820H	\$20.00
1.0	News	8:30P	ABC America	---M--	02/21/20	5:30P	ABC America This Morning	SEP TV021820H	\$20.00
1.0	News	8:30P	ABC America	---M--	02/21/20	5:30P	ABC America This Morning	SEP TV021820H	\$20.00
4.0	SPOT	10:14A	ABC Good Mon	---M--	02/19/20	7:30P	ABC Good Morning America	SEP TV021820H	\$90.00
4.0	SPOT	10:14A	ABC Good Mon	---M--	02/19/20	7:30P	ABC Good Morning America	SEP TV021820H	\$90.00
4.0	SPOT	10:14A	ABC Good Mon	---M--	02/20/20	7:30P	ABC Good Morning America	SEP TV021820H	\$90.00
4.0	SPOT	10:14A	ABC Good Mon	---M--	02/20/20	7:30P	ABC Good Morning America	SEP TV021820H	\$90.00
4.0	SPOT	10:14A	ABC Good Mon	---M--	02/21/20	7:30P	ABC Good Morning America	SEP TV021820H	\$90.00
4.0	SPOT	10:14A	ABC Good Mon	---M--	02/21/20	7:30P	ABC Good Morning America	SEP TV021820H	\$90.00
4.0	SPOT	10:14A	ABC Good Mon	---M--	02/22/20	7:30P	ABC Good Morning America	SEP TV021820H	\$90.00
4.0	SPOT	10:14A	ABC Good Mon	---M--	02/22/20	7:30P	ABC Good Morning America	SEP TV021820H	\$90.00
4.0	SPOT	10:14A	ABC Good Mon	---M--	02/23/20	7:30P	ABC Good Morning America	SEP TV021820H	\$90.00
11.0	SPOT	8:49A	ABC Good Mon	---M--	02/23/20	8:29AM	ABC Good Morning America Sunday	SEP TV021820H	\$40.00

???

Net 30 Payment is due within 30 days from invoice date.
Station owned or provided services by the Sinclair Broadcast Group.

Gross Total	\$760.00	Total Spots	13
Commission	(\$114.60)		
Net Total	\$645.40		

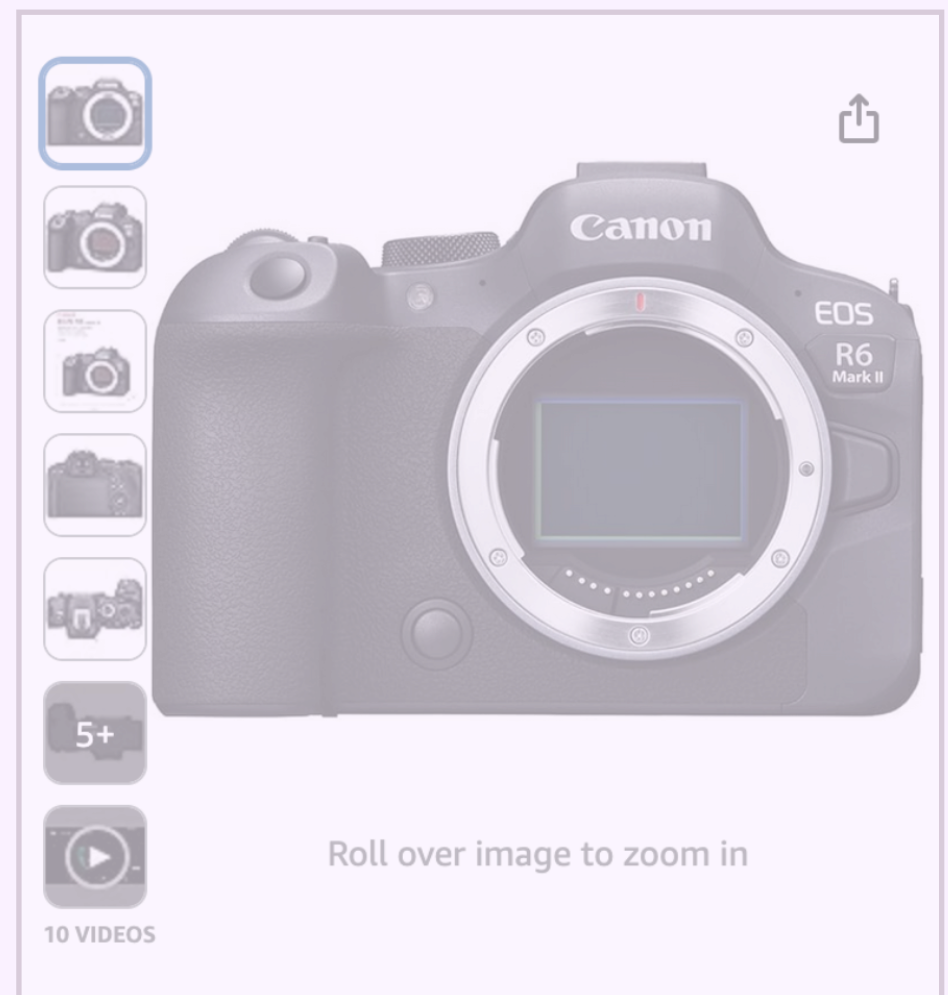
Officially printed on 02/27/20 11:44 AM Page 1

Hundreds or thousands of nodes

Unseen webpage

Electronics > Camera & Photo > Digital Cameras > Mirrorless Cameras

Canon EOS R6 Mark II Mirrorless Camera (Body Only), Full-Frame Camera, 24.2 Megapixel CMOS Sensor, Photo and Video Capabilities, Black



Visit the Canon Store

4.8 ★★★★★ (544)

Search this page

-20% \$1,999⁰⁰

List Price: \$2,499.00

Or \$79.11 /mo (48 mo). Select from 1 plan

FREE Returns



Ringer's scorers require structured documents.

~17 billion
OCR node
subsets

REMIT TO Sinclair Broadcast
45 WXLV
PO Box 206270
Dallas, TX 75320-6270

Advertiser: Citizens for Responsible Energy (84874)
Agency: Matsori Media LLC (19039)
Buyer: Media Assistant
Salesperson: Millennium DC, Washington DC (1108)
Special Handling: None
Product: POLITICAL ISSUE (1187)
Brand: NCHD29 (1340682)
Acct Types: National/Political/Issue/Agency BRD
Est/External#: 11637/ECR26803967
Demo:

Invoice: 7245066
Inv Date: 2/23/2020
Terms: CIA
Contract: 4248847
Bill Type: Weekly
Period: 2/23/2020 - 3/23/2020
CO-OP/Order Type: Normal
Package:
Gen Date: 2/25/2020 11:23:41 AM

Matsori Media LLC
1201 Hampton Street
Ste 3B
Columbia, SC 29201

Revisions: WXLV
Comments: WXLV

Line	Type	Scheduled	Schedule (Days to Run)	Air To	Length	Program	Copy/ESCI	Amount	Remarks
1.0	Contract Line								
1.0	News	8/3/20	ABC America	Mo --- -----	00:30	ABC America This Morning	SEP TV021820H	\$0.00	
1.0	News	8/3/20	ABC America	Tu --- -----	00:30	ABC America This Morning	SEP TV021820H	\$0.00	
1.0	News	8/3/20	ABC America	We --- -----	00:30	ABC America This Morning	SEP TV021820H	\$0.00	
1.0	News	8/3/20	ABC America	Th --- -----	00:30	ABC America This Morning	SEP TV021820H	\$0.00	
1.0	News	8/3/20	ABC America	Fr --- -----	00:30	ABC America This Morning	SEP TV021820H	\$0.00	
1.0	News	8/3/20	ABC America	Sa --- -----	00:30	ABC America This Morning	SEP TV021820H	\$0.00	
1.0	News	8/3/20	ABC America	Su --- -----	00:30	ABC America This Morning	SEP TV021820H	\$0.00	
1.0	SPO1	10/7/4	ABC Good Mon	Mo --- -----	00:30	ABC Good Morning America	SEP TV021820H	\$0.00	
1.0	SPO1	10/7/4	ABC Good Mon	Tu --- -----	00:30	ABC Good Morning America	SEP TV021820H	\$0.00	
1.0	SPO1	10/7/4	ABC Good Mon	We --- -----	00:30	ABC Good Morning America	SEP TV021820H	\$0.00	
1.0	SPO1	10/7/4	ABC Good Mon	Th --- -----	00:30	ABC Good Morning America	SEP TV021820H	\$0.00	
1.0	SPO1	10/7/4	ABC Good Mon	Fr --- -----	00:30	ABC Good Morning America	SEP TV021820H	\$0.00	
1.0	SPO1	10/7/4	ABC Good Mon	Sa --- -----	00:30	ABC Good Morning America	SEP TV021820H	\$0.00	
1.0	SPO1	10/7/4	ABC Good Mon	Su --- -----	00:30	ABC Good Morning America	SEP TV021820H	\$0.00	
1.0	SPO1	10/3/4	ABC Good Mon	Mo --- -----	00:30	ABC Good Morning America	SEP TV021820H	\$0.00	
1.0	SPO1	10/3/4	ABC Good Mon	Tu --- -----	00:30	ABC Good Morning America	SEP TV021820H	\$0.00	
1.0	SPO1	10/3/4	ABC Good Mon	We --- -----	00:30	ABC Good Morning America	SEP TV021820H	\$0.00	
1.0	SPO1	10/3/4	ABC Good Mon	Th --- -----	00:30	ABC Good Morning America	SEP TV021820H	\$0.00	
1.0	SPO1	10/3/4	ABC Good Mon	Fr --- -----	00:30	ABC Good Morning America	SEP TV021820H	\$0.00	
1.0	SPO1	10/3/4	ABC Good Mon	Sa --- -----	00:30	ABC Good Morning America	SEP TV021820H	\$0.00	
1.0	SPO1	10/3/4	ABC Good Mon	Su --- -----	00:30	ABC Good Morning America	SEP TV021820H	\$0.00	
Gross Total: \$760.00								Total Spots: 19	
Commission: \$114.00									
Net Total: \$646.00									

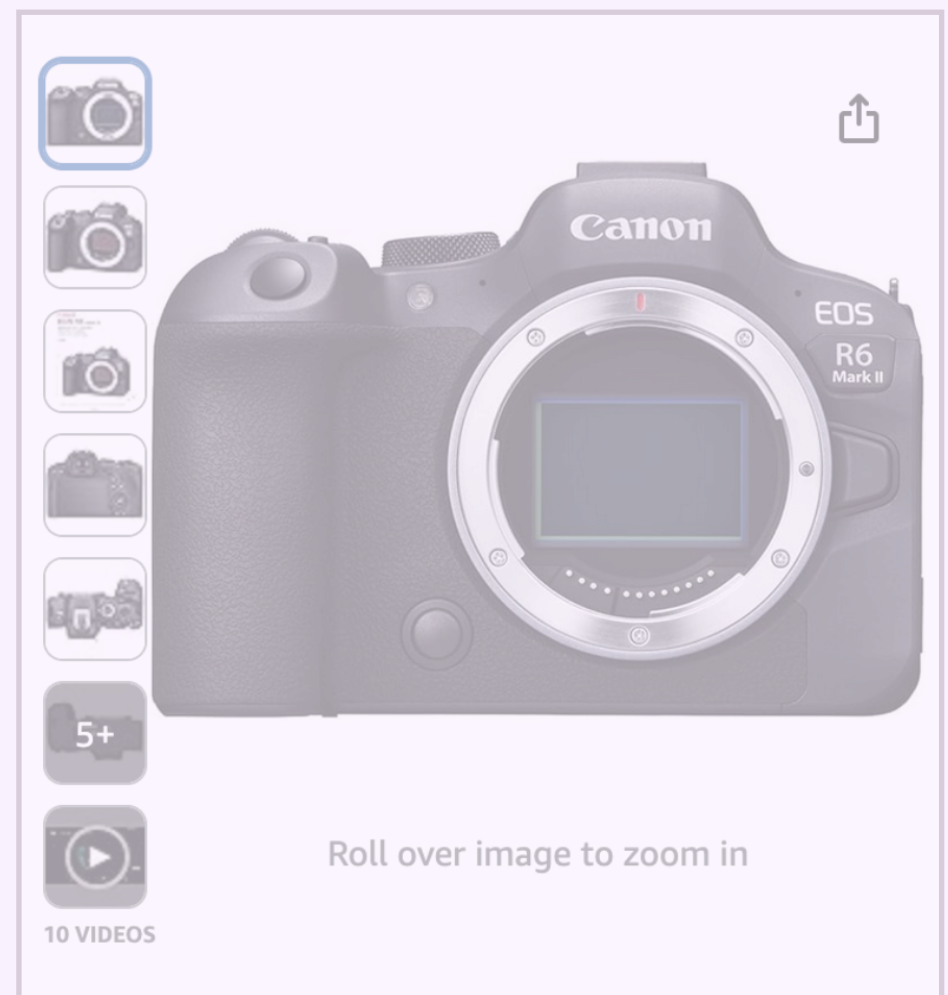
Net 30 Payment is due within 30 days from invoice date.
Station owned or provided services by the Sinclair Broadcast Group.

Officially printed on 02/27/20 11:44 AM

Unseen webpage

Electronics > Camera & Photo > Digital Cameras > Mirrorless Cameras

Canon EOS R6 Mark II Mirrorless Camera (Body Only), Full-Frame Camera, 24.2 Megapixel CMOS Sensor, Photo and Video Capabilities, Black



Visit the Canon Store

4.8 ★★★★★ (544)

Search this page

-20% \$1,999⁰⁰

List Price: \$2,499.00

Or \$79.11 /mo (48 mo). Select from 1 plan

FREE Returns



Ringer's scorers require structured documents.

~17 billion OCR node subsets

Hundreds or thousands of nodes

Projected Ringer benchmark completion time: the year 4105

The image shows a broadcast invoice from WXLV station. The header includes the station logo (abc 45 WXLV) and contact information for Matson Media LLC. A red box highlights the 'Salesperson' field: 'Millennium DC, Washington DC (1108) ph: (202) 955-5342 | fx: (202) 955-5348'. Another red box highlights the 'Advertiser' field: 'Citizens for Responsible Energy (84874)'. A third red box highlights the 'Product' field: 'POLITICAL ISSUE (75) (1187)'. The main body of the invoice is a table titled 'OFFICIAL BILLING INVOICE' with columns for Line, Type, Scheduled, Schedule (Days to Run), Air To, Length, Program, Copy/ESCI, Amount, and Remarks. The table lists multiple line items for news and spot programs. A red box highlights a portion of the table, and a red '???' is placed below it. The bottom of the invoice shows a total amount of \$760.00 and a net total of \$646.00. The footer includes 'Officially printed on 02/27/20 11:44AM' and 'Page 1'.

Unseen webpage

Electronics > Camera & Photo > Digital Cameras > Mirrorless Cameras

Canon EOS R6 Mark II Mirrorless Camera (Body Only), Full-Frame CMOS Sensor, Photo and Video Capabilities, Black

The image shows a product page for a Canon EOS R6 Mark II Mirrorless Camera. The main image is a large, high-quality photograph of the camera body in black. To the left of the main image are several smaller icons representing different views of the camera. Below the main image are several interactive elements: a '5+' icon, a '10 VIDEOS' icon, and a 'Roll over image to zoom in' prompt. The overall layout is clean and professional, typical of a retail website.

Visit the Canon Store

4.8 ★★★★★ (544)

Search this page

-20% \$1,999⁰⁰
List Price: \$2,499.00

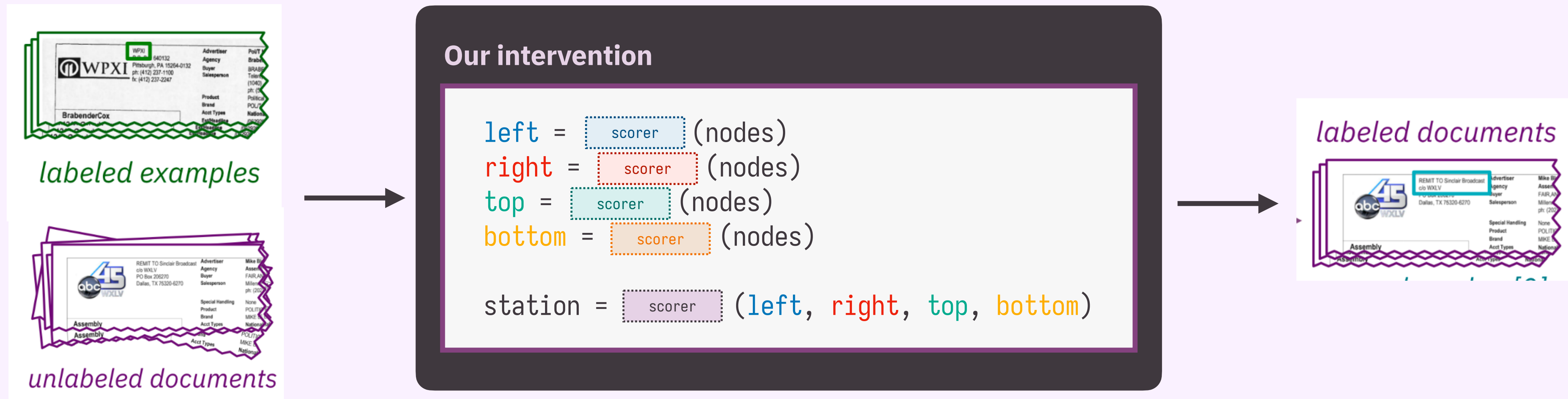
Or \$79.11/mo (48 mo). Select from 1 plan

FREE Returns



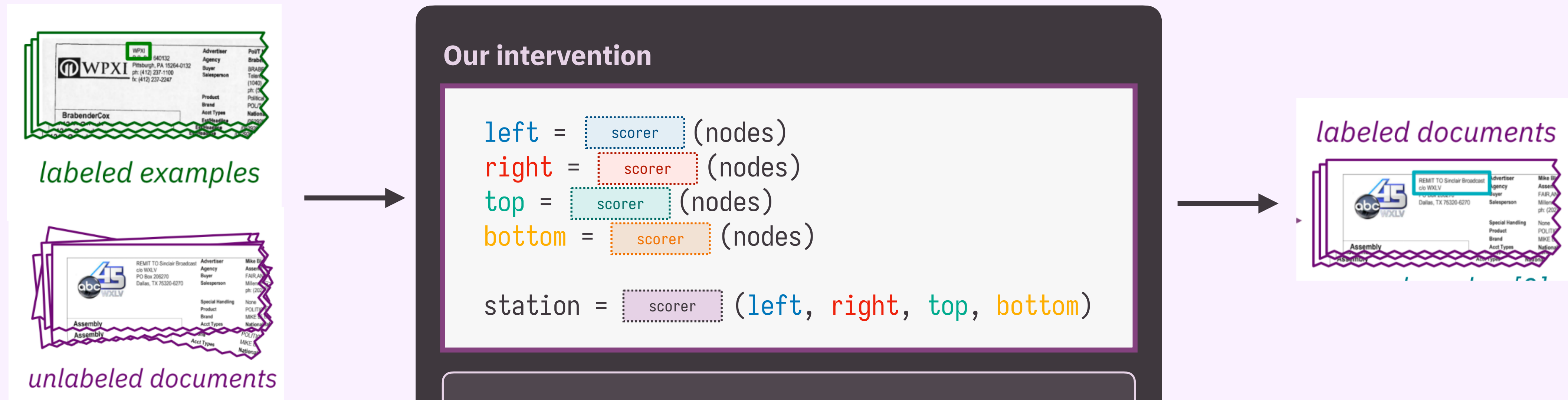
Ringer's scorers require structured documents.

Program sketches with scorer holes adapt traditional synthesis for messy data.



Ringer's scorers require structured documents.

Program sketches with scorer holes adapt traditional synthesis for messy data.



Program sketches with scorer holes adapt traditional synthesis for messy data.
 In particular, we can synthesize multiple scorers to identify information!

FOX 4
K B T V

REMIT TO Sindair Broadcast Gro
c/o KBTV
PO Box 206270
Dallas, TX 75320-6270

Advertiser
Agency
Buyer
Salesperson

Special Handling
Product
Brand
Acct Types
Est/External #
Demo
Revision
Comments

AKPD Message & Media
730 N Franklin
Ste 404
Chicago, IL 60654
AgM

Beaumont (KBTV) **OFFICIAL BILL**

Line	Type	Scheduled	Schedule Days to Run	Air Time
1.0	Contract Line Remark:		<u>7a news</u>	

Program sketches with scorer holes adapt traditional synthesis for messy data.
In particular, we can synthesize multiple scorers to identify information!

```

left = scorer (nodes)
right = scorer (nodes)
top = scorer (nodes)
bottom = scorer (nodes)

station = scorer (left, right, top, bottom)

```

Synthesize scorers to find the edges of the field, then use those results to find the field itself.



REMIT TO Sinclair Broadcast Gro
c/o KBTU

PO Box 206270
Dallas, TX 75320-6270

Advertiser
Agency
Buyer
Salesperson

Special Handling
Product
Brand
Acct Types
Est/External #
Demo
Revision
Comments

AKPD Message & Media
730 N Franklin
Ste 404
Chicago, IL 60654

AgM

Beaumont (KBTU)

OFFICIAL BILL

Line	Type	Scheduled	Schedule Days to Run	Air Time
1.0	Contract Line Remark:		<u>7a news</u>	

Program sketches with scorer holes adapt traditional synthesis for messy data.
In particular, we can synthesize multiple scorers to identify information!

Program sketch with scorer holes

```

left = scorer (nodes)
right = scorer (nodes)
top = scorer (nodes)
bottom = scorer (nodes)

station = scorer (left, right, top, bottom)

```

```

left = scorer_l(nodes)
right = scorer_r(nodes)
top = scorer_t(nodes)
bottom = scorer_b(nodes)

station = scorer_s(left, right, top, bottom)

```

Concrete scorer program

Marshall
(our tool)



REMIT TO Sinclair Broadcast Gro
c/o KBTV
PO Box 206270
Dallas, TX 75320-6270

Advertiser
Agency
Buyer
Salesperson

Special Handling
Product
Brand
Acct Types
Est/External #
Demo
Revision
Comments

AKPD Message & Media
730 N Franklin
Ste 404
Chicago, IL 60654

AgM

Beaumont (KBTV)

OFFICIAL BILL

Line	Type	Scheduled	Schedule Days to Run	Air Time
1.0	Contract Line Remark:		<u>7a news</u>	

Program sketches with scorer holes adapt **sketch synthesis** for messy data.

In particular, we can synthesize multiple scorers to identify information!

Program sketch with scorer holes

```

left = scorer (nodes)
right = scorer (nodes)
top = scorer (nodes)
bottom = scorer (nodes)

station = scorer (left, right, top, bottom)

```

```

left = scorer_l(nodes)
right = scorer_r(nodes)
top = scorer_t(nodes)
bottom = scorer_b(nodes)

station = scorer_s(left, right, top, bottom)

```

Concrete scorer program

Marshall
(our tool)

sketch synthesizer

Program sketch

```

def two_times(num):
    result = num << ??
    return result

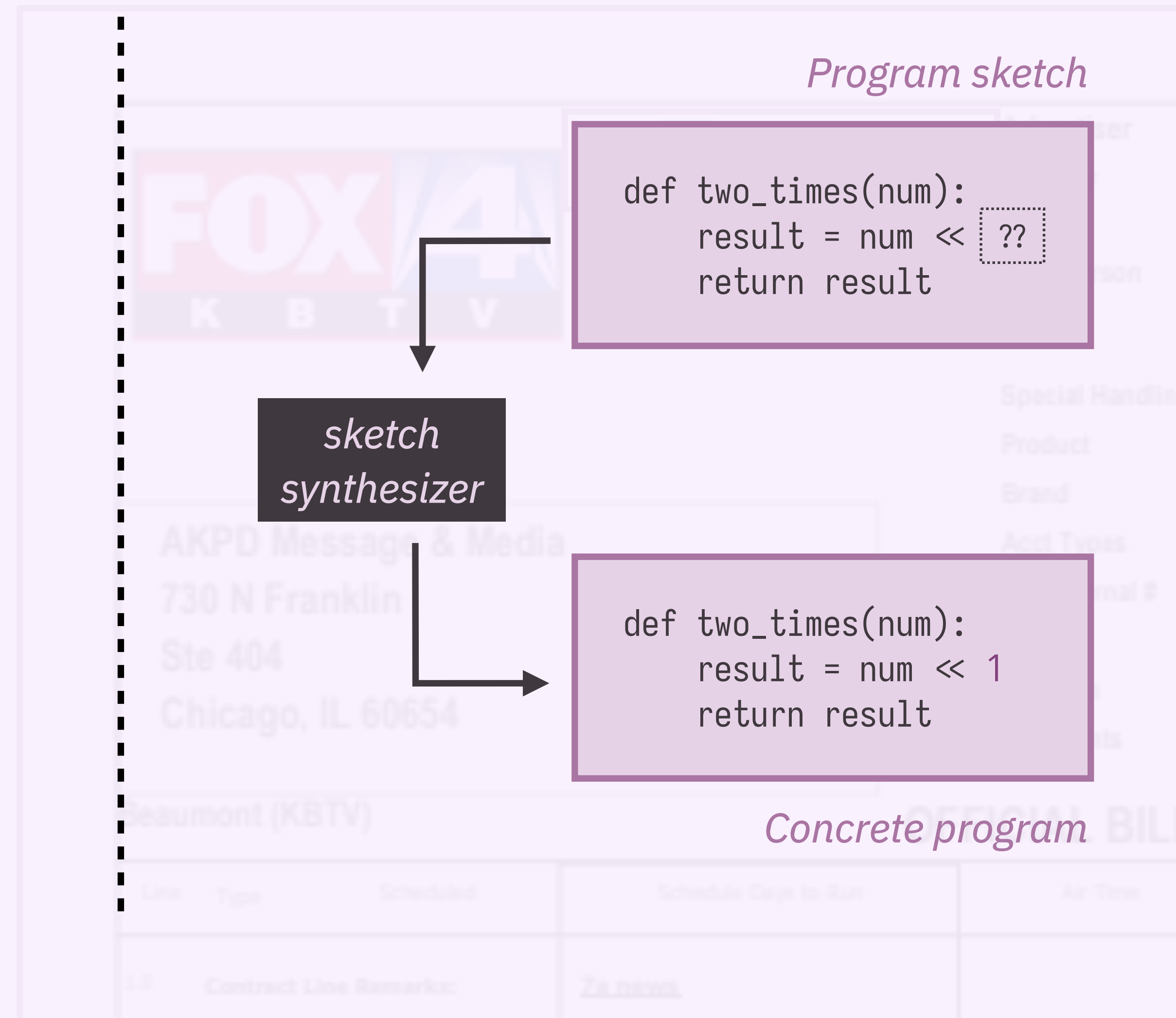
```

```

def two_times(num):
    result = num << 1
    return result

```

Concrete program



Program sketches with scorer holes adapt traditional synthesis for messy data.
In particular, we can synthesize multiple scorers to identify information!

```

left = scorer (nodes)
right = scorer (nodes)
top = scorer (nodes)
bottom = scorer (nodes)

station = scorer (left, right, top, bottom)

```

```

inner = scorer (nodes)
station = scorer (inner)

```

Synthesize one scorer to rank OCR boxes by likelihood of field inclusion, then synthesize another scorer to find the field itself.

REMIT TO Sinclair Broadcast Gro
c/o KBTU
PO Box 206270
Dallas, TX 75320-6270

Advertiser
Agency
Buyer
Salesperson

Special Handling
Product
Brand
Acct Types
Est/External #
Demo
Revision
Comments

AKPD Message & Media
730 N Franklin
Ste 404
Chicago, IL 60654
AgM

Beaumont (KBTU) **OFFICIAL BILL**

Line	Type	Scheduled	Schedule Days to Run	Air Time
1.0	Contract Line Remark:		7a news	

Program sketches with scorer holes adapt traditional synthesis for messy data.

Different program sketches
with scorer holes...



...yield different extraction
tools for many domains.

```

left = scorer (nodes)
right = scorer (nodes)
top = scorer (nodes)
bottom = scorer (nodes)
station = scorer (left, right, top, bottom)
  
```

```

inner = scorer (nodes)
station = scorer (inner)
  
```

Line	Type	Schedule	Schedule Day to Run	Air Time	Length	Copy/Sec	Amount	Remarks
1.0	Contract Line Remark:	Za...						
	News	3/23/22 News The News	Mo-Fri	3/23/22 7:35 AM (M)	00:30	PPA 20 0300	\$55.00	
2.0	Contract Line Remark:	Ma...						
	News	3/23/22 News The News	Mo-Fri	3/23/22 8:25 AM (M)	00:30	PPA 20 0300	\$55.00	

Tools for *PDFs*


```

landmark = scorer (nodes)
target = scorer (landmark)
  
```

```

DOM node = scorer (nodes)
  
```

Electronics › Camera & Photo › Digital Cameras › Mirrorless Cameras



Nikon Z6 III | Full-Frame mirrorless Stills/Video Camera with 6K/60p Internal RAW Recording | Nikon USA Model

Visit the Nikon Store
4.4 ★★★★★ (59) | Search this page
100+ bought in past month

\$2,496⁹⁵

Or **\$98.82**/mo (48 mo). Select from 1 plan
FREE Returns
Save up to 2% with business pricing. Sign up for a free Amazon Business account

Pay **\$208.08/month for 12 months, interest-free upon approval for Amazon Visa**

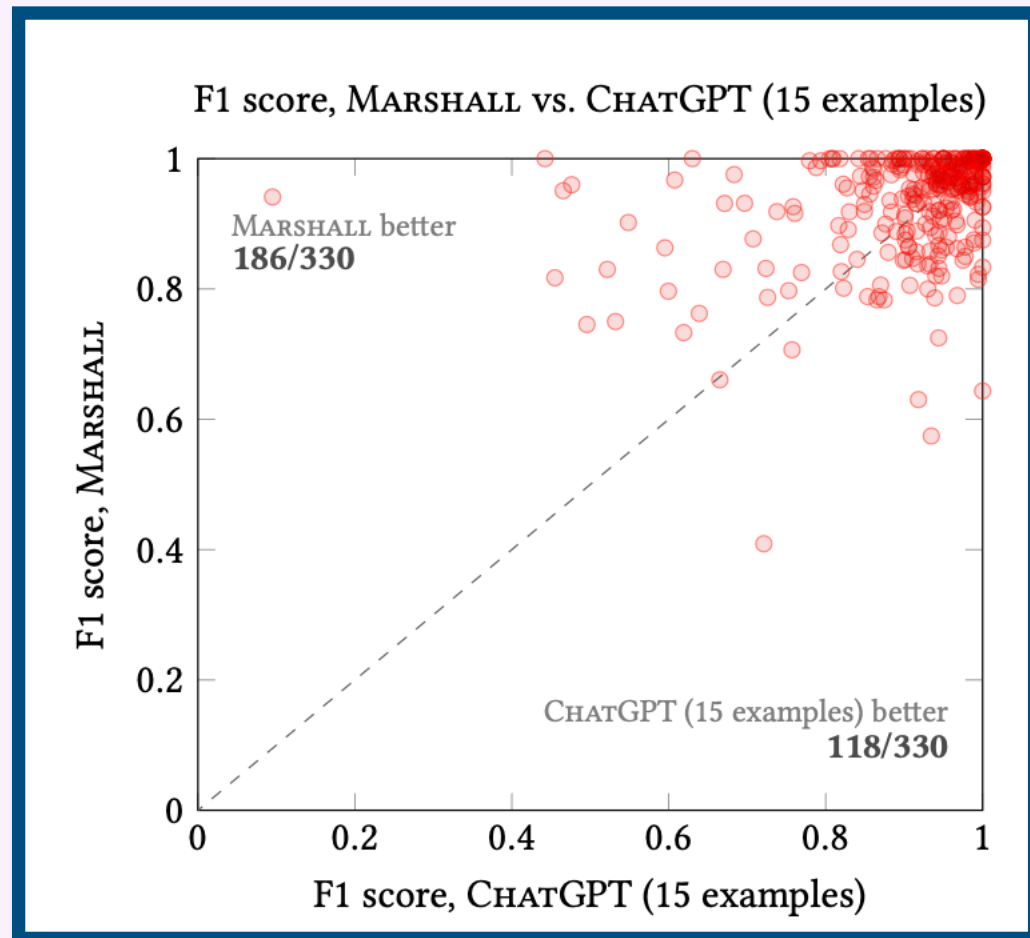
Tools for *webpages*



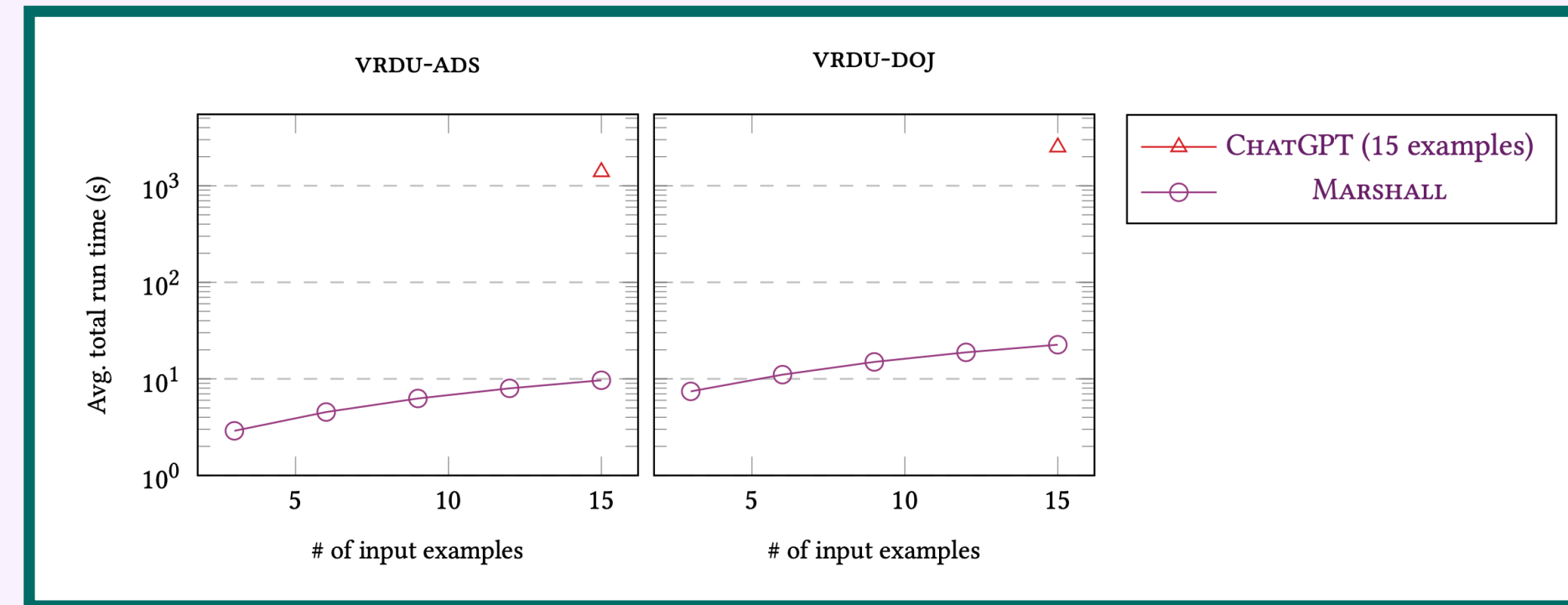
Does this work for actual datasets?

Does this work for actual datasets?

versus LLMs



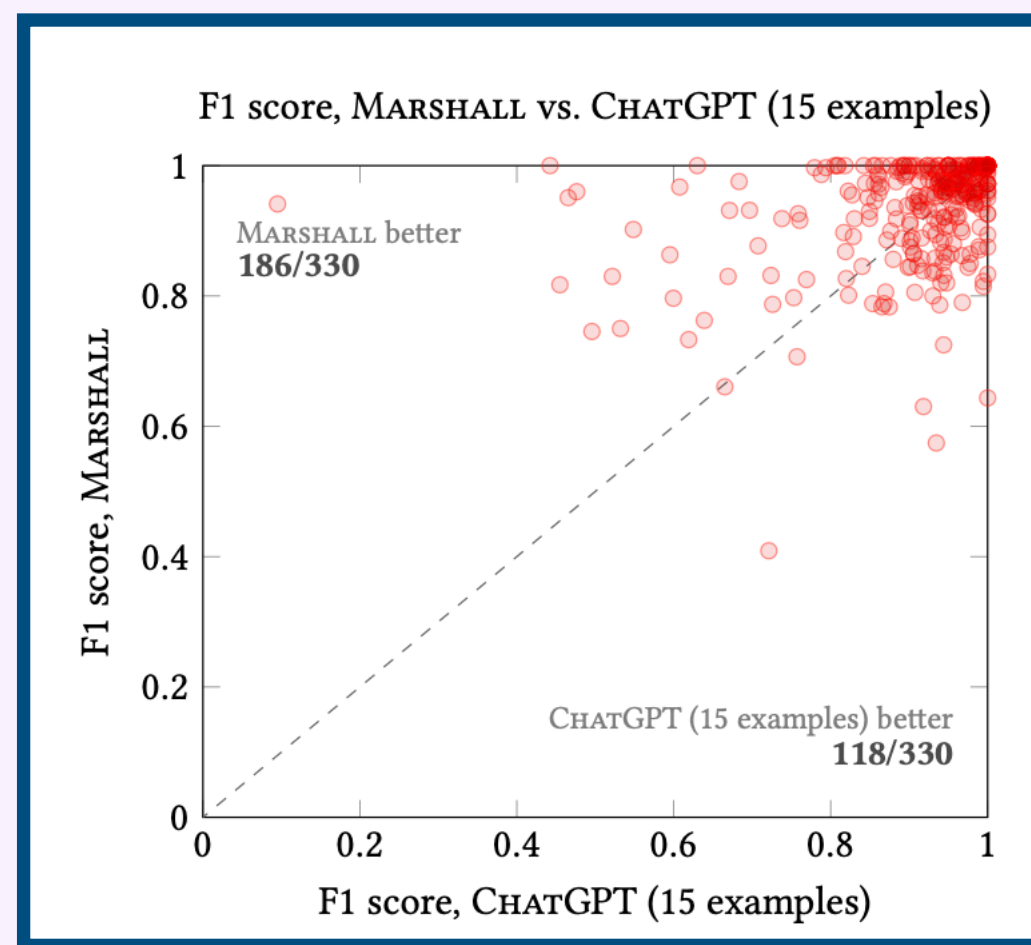
Our approach achieves precision and accuracy parity with GPT-4o...



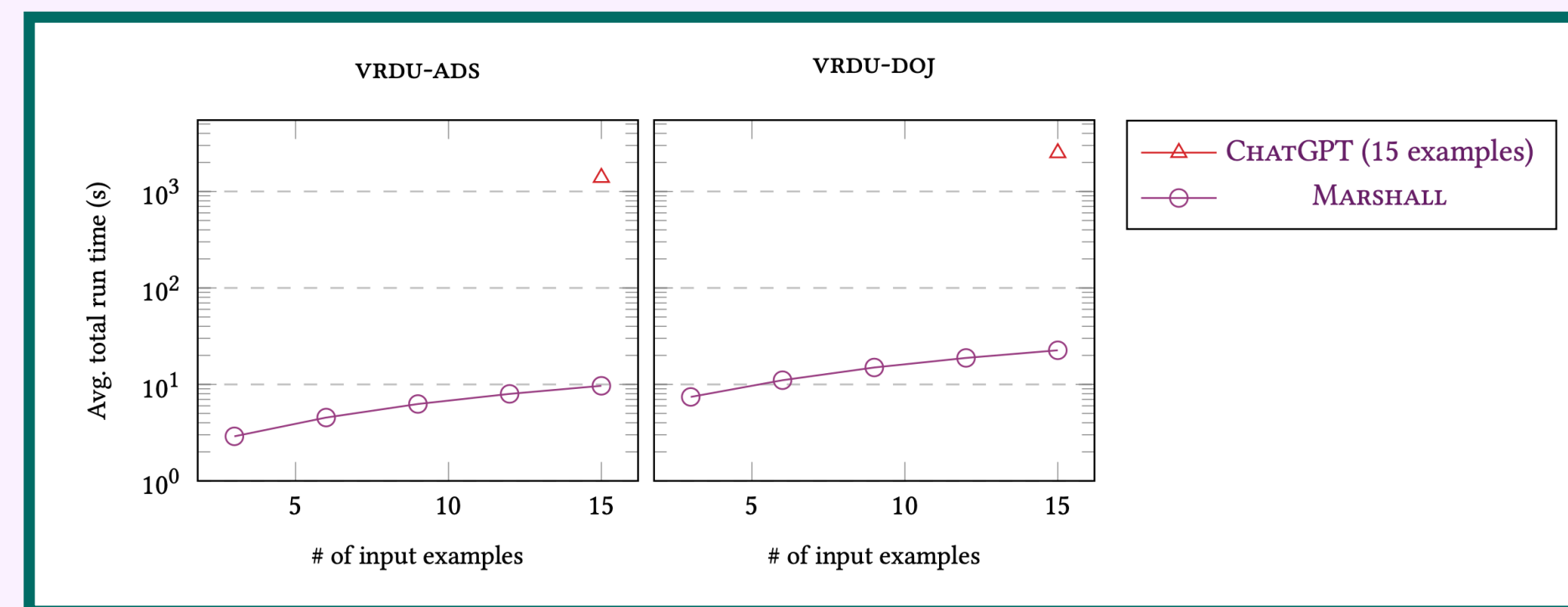
while running $\approx 120\times$ faster and $\approx 95,700\times$ cheaper.

Does this work for actual datasets?

versus LLMs



Our approach achieves precision and accuracy parity with GPT-4o...



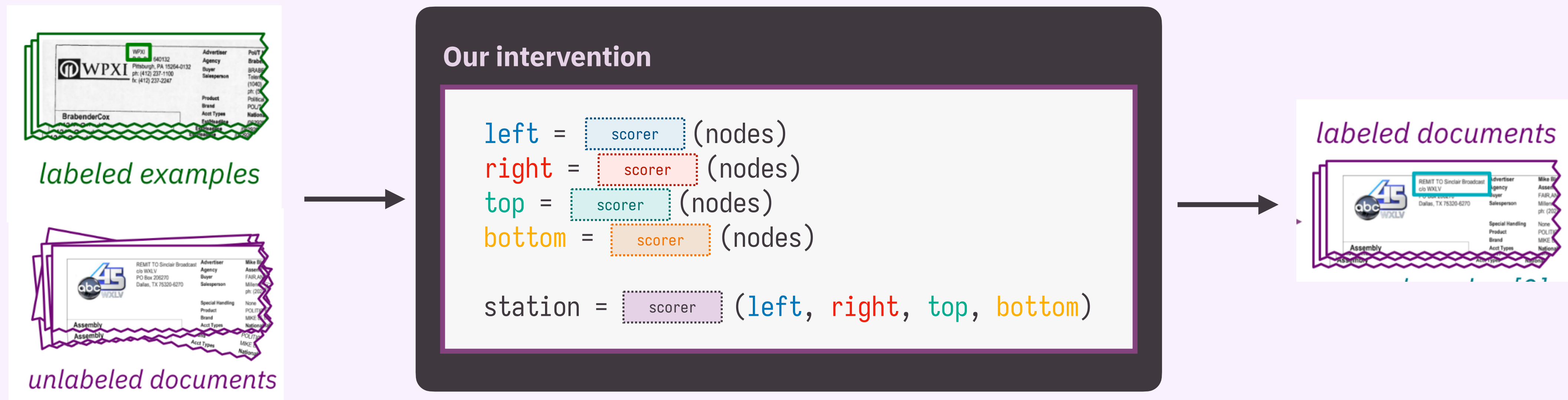
while running $\approx 120\times$ faster and $\approx 95,700\times$ cheaper.

versus traditional synthesis

MARSHALL is 4,216,020.35 \times faster than RINGER.

Program Sketches with Scorer Holes

an approach for retrieving info from heterogenous data by synthesizing multiple scorers.



David Minh-Duy Cao, Sarah E. Chasins
dmcao@berkeley.edu