

Human-Centered Tools for Reliable Use of Machine Translation

Samantha Robertson
Ph.D. Student, EECS

EPIC Advance
October 26, 2022



Hold the kidney
medicine until
you have a
chance to speak
with your kidney
doctor



Hold the kidney
medicine until
you have a
chance to speak
with your kidney
doctor



English



Mandarin

[Khoong et al., 2019]



English

Hold the kidney
medicine until
you have a
chance to speak
with your kidney
doctor



Mandarin

[Khoong et al., 2019]



English

Hold the kidney medicine until you have a chance to speak with your kidney doctor



继续服用肾脏药物直到您有机会和您的肾脏医生交流



Mandarin

[Khoong et al., 2019]

Do you give this translation to the patient?

Hold the kidney medicine until you have a chance to speak with your kidney doctor



继续服用肾脏药物直到您有机会和您的肾脏医生交流



English



Mandarin

[Khoong et al., 2019]

Potentially life-threatening error

Hold the kidney medicine until you have a chance to speak with your kidney doctor



继续服用肾脏药物直到您有机会和您的肾脏医生交流

Keep taking kidney medicine until you talk to your kidney doctor



English



Mandarin

[Khoong et al., 2019]

Hard to identify errors

Hold the kidney medicine until you have a chance to speak with your kidney doctor



English



继续服用肾脏药物直到您有机会和您的肾脏医生交流

Keep taking kidney medicine until you talk to your kidney doctor



Mandarin

[Khoong et al., 2019]

Lack of actionable strategies to recover

Hold the kidney medicine until you have a chance to speak with your kidney doctor



继续服用肾脏药物直到您有机会和您的肾脏医生交流



Keep taking kidney medicine until you talk to your kidney doctor



English



Mandarin

[Khoong et al., 2019]

There are lots of “translation” problems in EPIC

Check out Nithin’s poster for more!



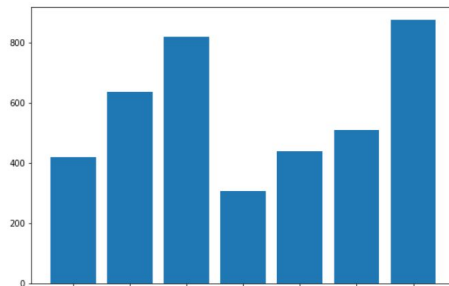
What are the demographics of traffic stop subjects?



```
stops_df = df[df.incident_type == "traffic stop"]
ax = stops_df['race']\
      .value_counts()\
      .plot(kind='bar')
```



Natural language



Code and/or visualization



MT has the potential to massively improve communication across language barriers,

especially for people who live and work in places where they don't speak the dominant language,

but it doesn't always work.

Errors can have *high-stakes consequences*

8% of sentences from emergency room discharge instructions translated by Google Translate to Chinese had **potential for clinically significant harm**.

[Khoong, et al., 2019]

17.5% of translations of emergency discharge instructions across seven languages **did not retain the meaning**, with worst accuracy for Farsi (**32.5%**), and Armenian (**45%**)

[Taira, et al., 2021]

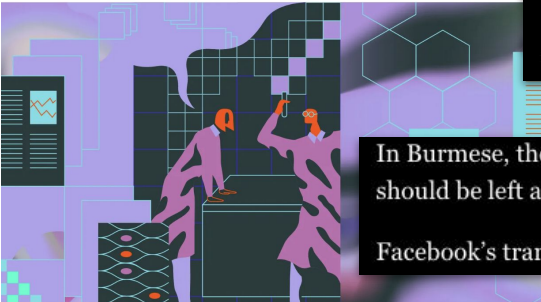
Errors can have *high-stakes consequences*

Google Translate still isn't good enough for medical instructions

For some languages, it can give patients misleading information

By Nicole Wetzman | Mar 9, 2021, 8:40am EST

f t SHARE



Why Facebook is losing the war on hate speech in Myanmar

By STEVE STECKLOW | Filed Aug. 15, 2018, 3 p.m. GMT

In Burmese, the post says: "Kill all the kalars that you see in Myanmar; none of them should be left alive."

Facebook's translation into English: "I shouldn't have a rainbow in Myanmar."

"I have lost work— I have lost good clients because they don't have the patience to communicate with me. One of the clients told me that using translator— you work very well, but I need someone that I can communicate with."

[Liebling, Lahav, et al., 2020]

We need tools that help users

identify and

recover from

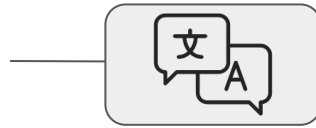
translation errors



- Translation as a specification problem
- What do users need?
- Aside: Tools for professional translators
- Reframing the problem: from specification to retrieval
- Takeaways

- **Translation as a specification problem**
- What do users need?
- Aside: Tools for professional translators
- Reframing the problem: from specification to retrieval
- Takeaways

Your child is fitting



(In Swahili:)
Your child is dead

Provide more information
about the quality of *outputs*

Your child is fitting



(In Swahili:)
Your child is dead



- Extremely difficult!

Suggest better
inputs

Provide more information
about the quality of *outputs*

Your child is fitting



(In Swahili:)
Your child is dead



- Extremely difficult!

Suggest better
inputs

Provide more information
about the quality of *outputs*

Your child is fitting



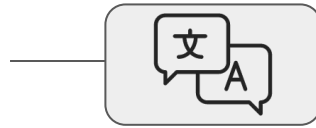
(In Swahili:
Your child is dead



- Still difficult
- Responsibility on the user
- No guarantees

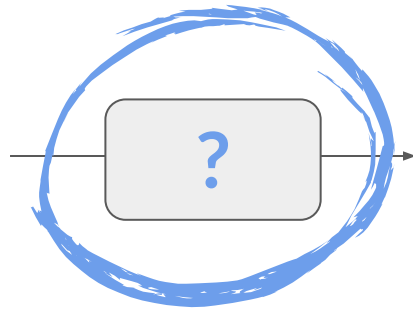
- Extremely difficult!

Your child is fitting



(In Swahili:)
Your child is dead

Your child is fitting



(In Swahili:)

Your child is dead

- Translation as a specification problem
- **What do users need?**
- Aside: Tools for professional translators
- Reframing the problem: from specification to retrieval
- Takeaways

Taking translation *models*
out of the picture,

interpretation quality is still
a challenge

Clinician strategies

- Teach back
- Back-translation
- Non-verbal communication
- Seeking **verification**
- **Pragmatic** approach

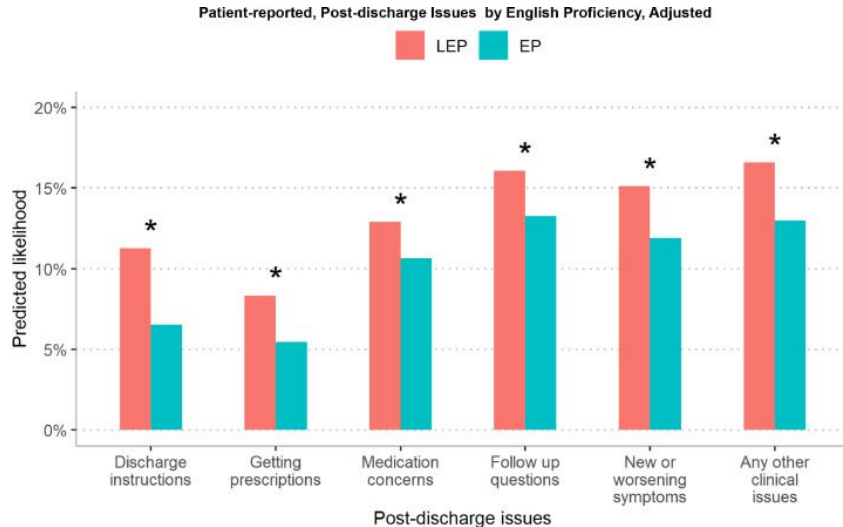
Consistent with preference for **phrase-based** tools

[Panayiotou, 2019, 2020; Spechbach, 2017; Turner, 2019]

Building
verifiable & pragmatic
translation tools
(for hospital discharge instructions)

Discharge instructions

- Often not translated
- Highly structured & repetitive



Dear __,

You were seen in the UCSF emergency room for ____.

We did __ tests and found that you have __.

We have prescribed you __, which you should take __. Please stop taking __ and continue your other medications.

If you experience __ please call your primary care doctor.

“Patients with LEP were more likely to report a problem with all measured postdischarge issues”

[Malevanchik, et al., 2021]

- Translation as a specification problem
- What do users need?
- **Aside: Tools for professional translators**
- Reframing the problem: from specification to retrieval
- Takeaways

Computer-Aided Translation & Translation Memories

Part of the segment has been translated before

The screenshot displays a CAT tool window titled "Fragment Matches - SDL Marketing New_EN-US_FR". The main area shows a match between an English source segment and a French translation segment. The source segment is highlighted in yellow and reads: "As Gerard Schram, Web Strategy and Operations Manager explains, 'Ultimately, a poor customer experience leads to frustration and impacts loyalty – that's what we were facing with our platform.'" The translation segment is highlighted in blue and reads: "Gerard Schram, responsable stratégie et opérations Web". A black arrow points from the text "Part of the segment has been translated before" to the blue translation segment. Below this, another window titled "SKF case study 2.docx.sdx.kliff [Translation]" is visible, showing a table with two rows. The first row contains the same English source text and a French translation: "300 000 produits.". A black arrow points from the text "Writing a new translation" to this row. The second row contains a different English source text and a blank translation field.

28	As Gerard Schram, Web Strategy and Operations Manager explains, "Ultimately, a poor customer experience leads to frustration and impacts loyalty – that's what we were facing with our platform."	300 000 produits.
29	We wanted to present one face to the customer and make it easy for them to find everything they need about SKF on a single, consistently branded site while enjoying a local experience."	
	Compounding matters, a decentralized	

Writing a new translation

Computer-Aided Translation & Translation Memories

Fragment Matches - SDL Marketing New_EN-US_FR

As [Gerard Schram, Web Strategy and Operations Manager](#), explains, "Ultimately, a poor customer experience leads to frustration and impacts

1 **Gerard Schram, Web Strategy and Operations Manager**

SDL Marketing New_EN-US_FR

Concordance Search Translation

SKF case_study 2.docx.sdlxliff [Translation]

28 As Gerard Schram, Web Strategy and Operations Manager, explains, "Ultimately, a poor customer experience leads to frustration and impacts – that's what we were facing with We wanted to present one face and make it easy for them to find need about SKF on a single, compact site while enjoying a local experience."

29 Compounding matters, a decentralized

1 The hares and the tortoise

80% El liebre y la tortuga

The hare hares and the tortoise

Close TM diff (ESC)

UNREVIEWED • Example User (example@lilt.com) • 2021-10-25 08:24

Source: Lilt

Source: Trados

“Fuzzy” match

Computer-Aided Translation & Translation Memories

The screenshot shows a CAT tool window titled 'Fragment Matches - SDL Marketing New_EN-US_FR'. The main text area displays a match for the sentence 'The hares and the tortoise' with an 80% match rate. The source text is 'The hare hares and the tortoise' and the target text is 'El liebre y la tortuga'. The match is marked as 'UNREVIEWED' and is attributed to 'Example User (example@lilt.com)' on '2021-10-25 08:24'. The interface also shows a list of other matches on the left and a 'P' button at the bottom right.

Designed specifically for professional translators

- speed
- consistency

Source: Lilt

Source: Trados

Discharge instructions

- Often not translated
- Highly structured & repetitive

Imagine we have a
translation memory
for discharge instructions

Dear __,

You were seen in the UCSF emergency room
for ____.

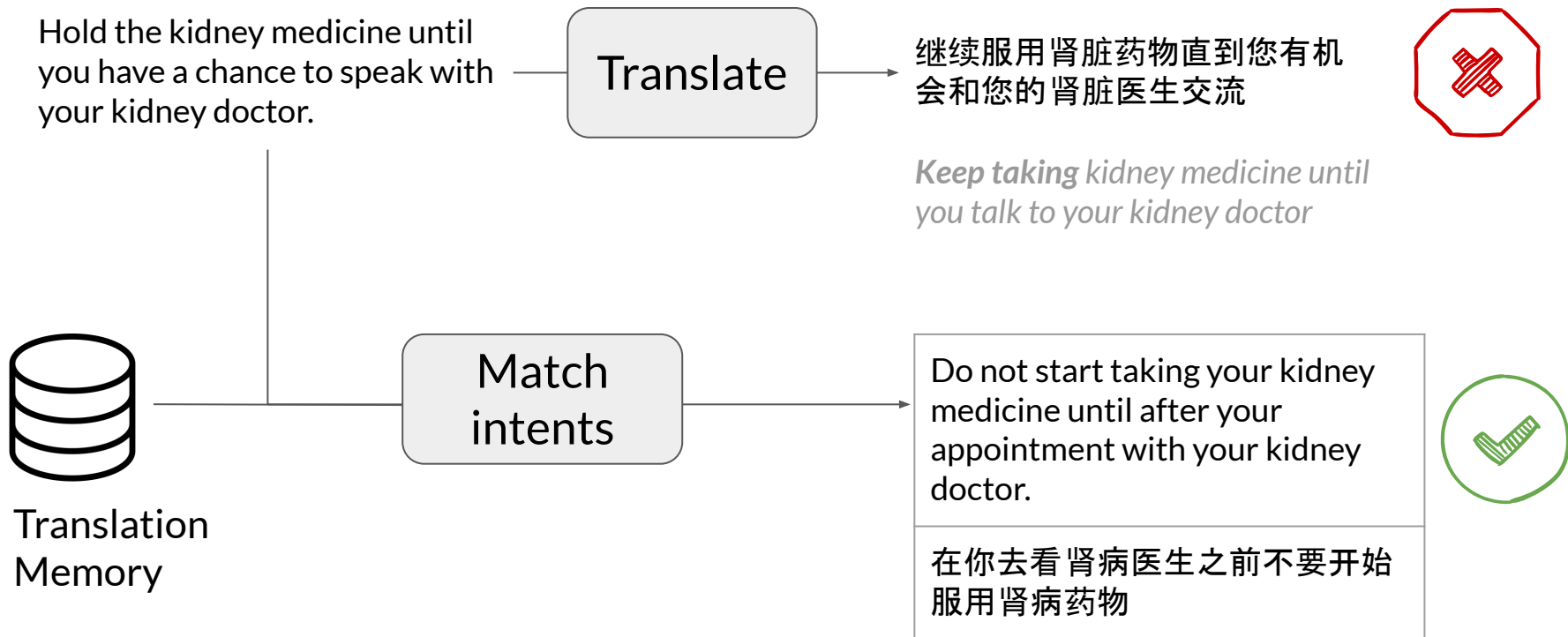
We did __ tests and found that you have __.

We have prescribed you __, which you should
take __. Please stop taking __ and continue
your other medications.

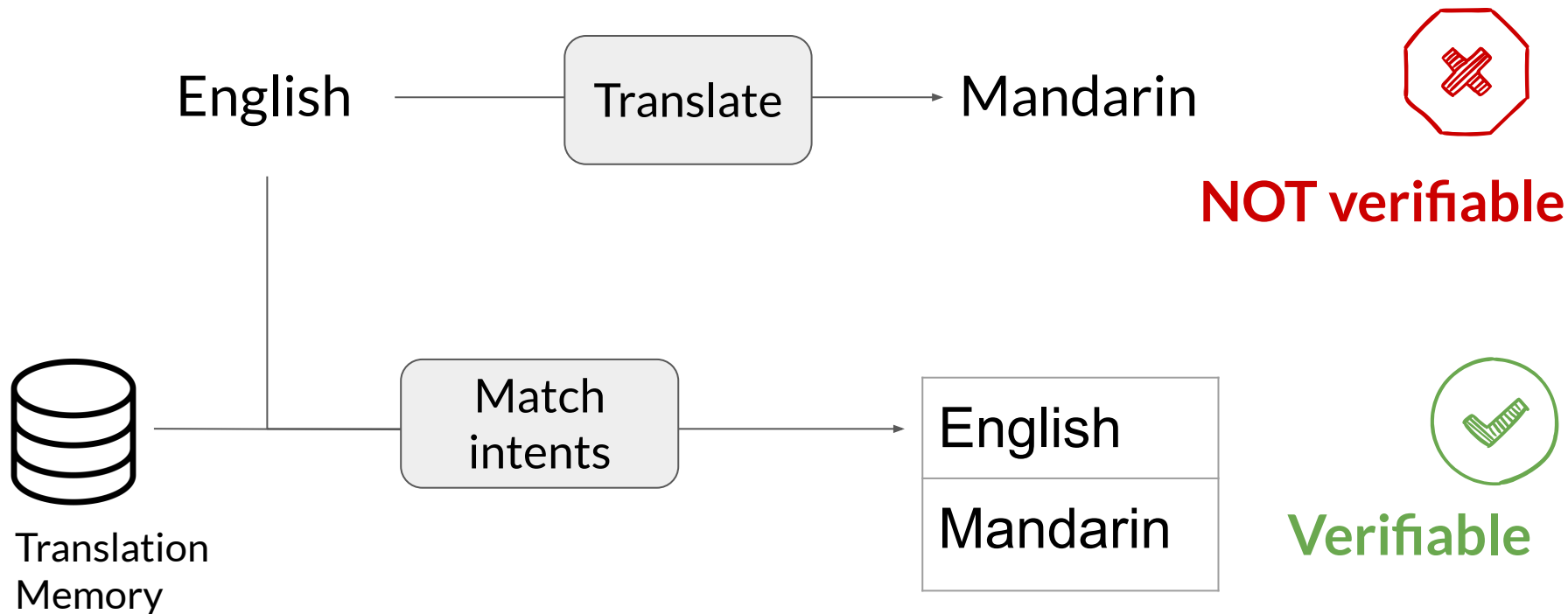
If you experience __ please call your primary
care doctor.

- Translation as a specification problem
- What do users need?
- Aside: Tools for professional translators
- **Reframing the problem: from specification to retrieval**
- Takeaways

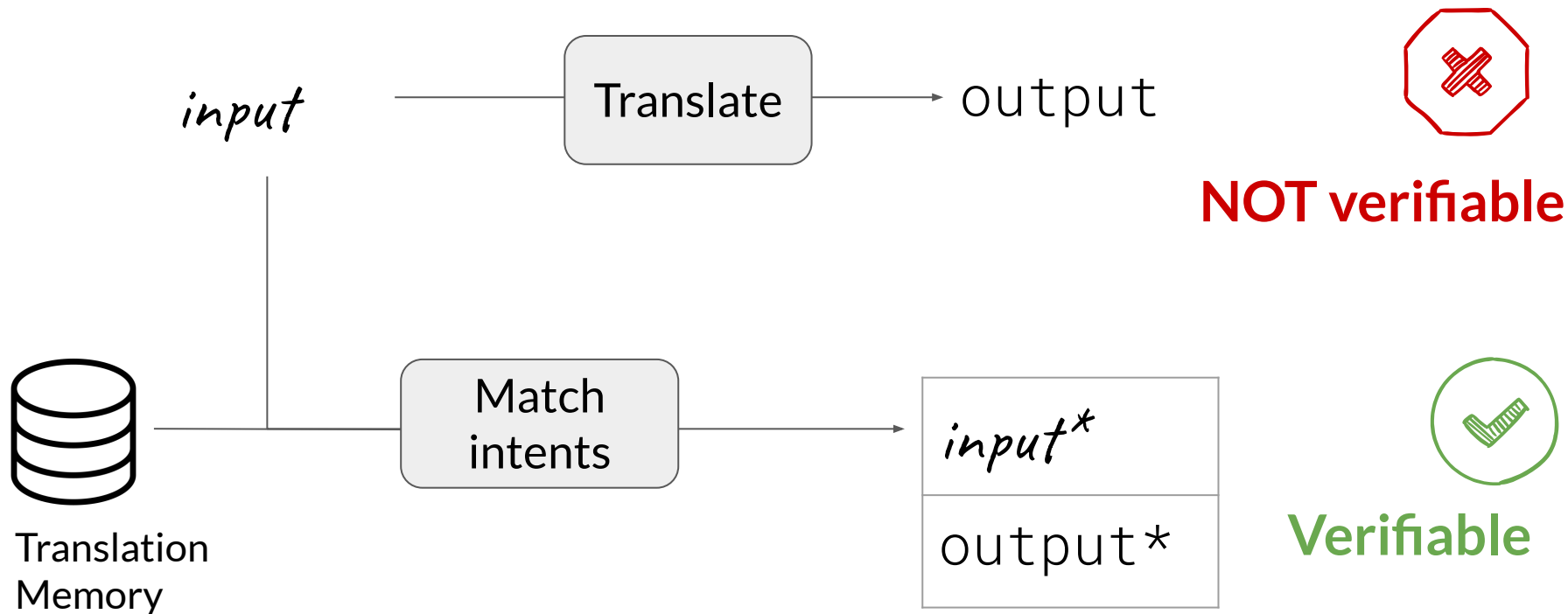
Translation becomes a retrieval problem with **verifiable** output



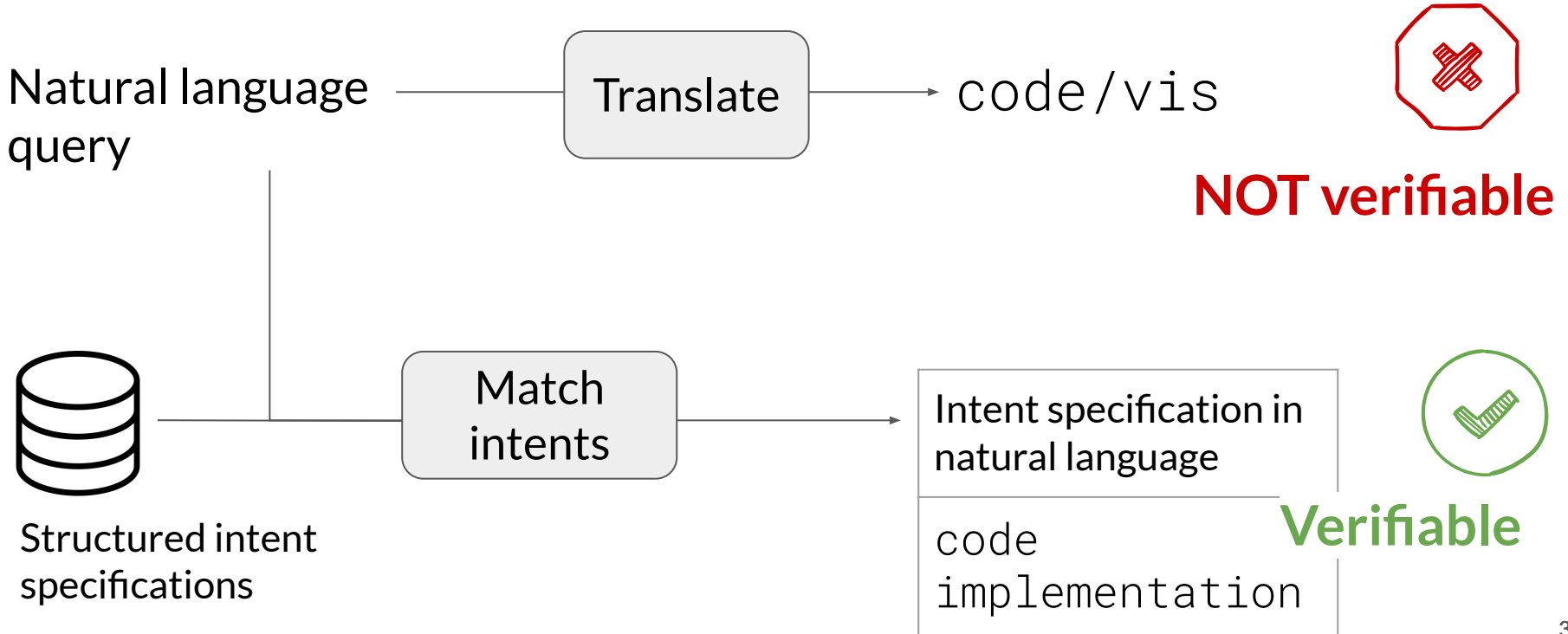
Translation becomes a retrieval problem with **verifiable** output



Translation becomes a retrieval problem with **verifiable** output



Translation becomes a retrieval problem with **verifiable** output

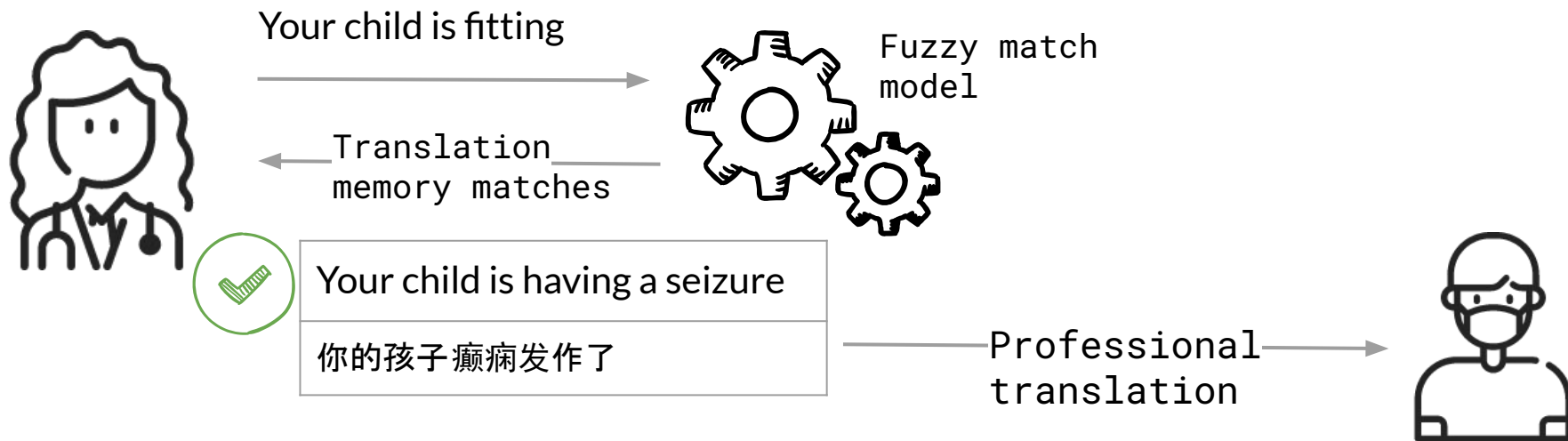


Developing custom translation memory

3,000+ emergency department discharge instructions from UCSF

TOPIC		Count		
▼	Admitted for...	Count 12		
13	You were admitted for <SYMPTOM/CONDITION>.		Admitted for...	
14	You were admitted for concern for <CONDITION> or <CONDITION>.		Admitted for...	
15	You were admitted because you had <SYMPTOM>.		Admitted for...	
16	You were admitted for <SYMPTOMS> thought to be due to <CONDITION>.		Admitted for...	
17	You were admitted to the hospital for <CONDITION/SYMPTOM>.		Admitted for...	
18	You were admitted for several symptoms including <SYMPTOM>.		Admitted for...	
19	You were admitted for <SYMPTOM> which we believe is from <CONDITION>.		Admitted for...	
20	You were seen and evaluated in the <DOCTOR/OFFICE> today for <SYMPTOM/CONDITION>.		Admitted for...	
21	You were seen in the ED for <SYMPTOM/CONDITION>.		Admitted for...	
22	You have been seen today for <SYMPTOMS> that have been going on for <TIMEFRAME>.		Admitted for...	
23	You were evaluated here at UCSF for your <SYMPTOMS>.		Admitted for...	
24	You were seen in the <DOCTOR/OFFICE> for <BODY PART> pain.		Admitted for...	
+				
TOPIC		Count		
▼	Tests	Count 43		
25	Your blood pressure was <a little> <high/low>.		Tests	
26	We found that you are <CONDITION> but otherwise your <TEST> was normal.		Tests	
27	Your <TESTS> showed <CONDITIONS/SYMPTOMS>.		Tests	
28	Your <TESTS> showed <TEST RESULT> which means you might have <CONDITION>.		Tests	
29	Your <TEST> was normal.		Tests	
30	Your <TEST> showed <no> <TEST RESULT/CONDITION>.		Tests	
31	Your <TEST> do not show any <TEST RESULT/CONDITION>.		Tests	
32	Your <TEST> did not show signs of <TEST RESULT/CONDITION>.		Tests	

Using the translation memory

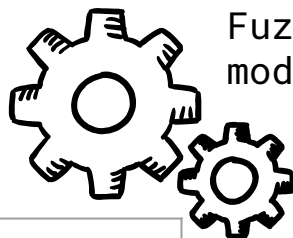


Using the translation memory

Hold the kidney medicine until you have a chance to speak with your kidney doctor.



Translation
memory matches



Fuzzy match
model

Do not take the kidney medicine.

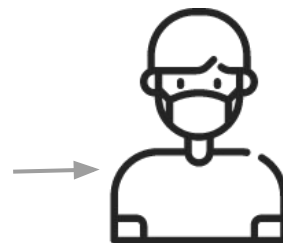
不要服用你的肾病药物

+

Please see your kidney doctor soon

请和你的肾病医生预约下次诊疗

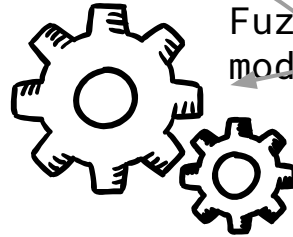
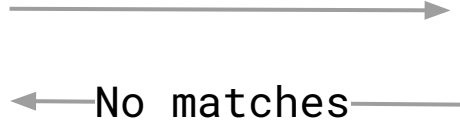
MT with
constraints



Using the translation memory



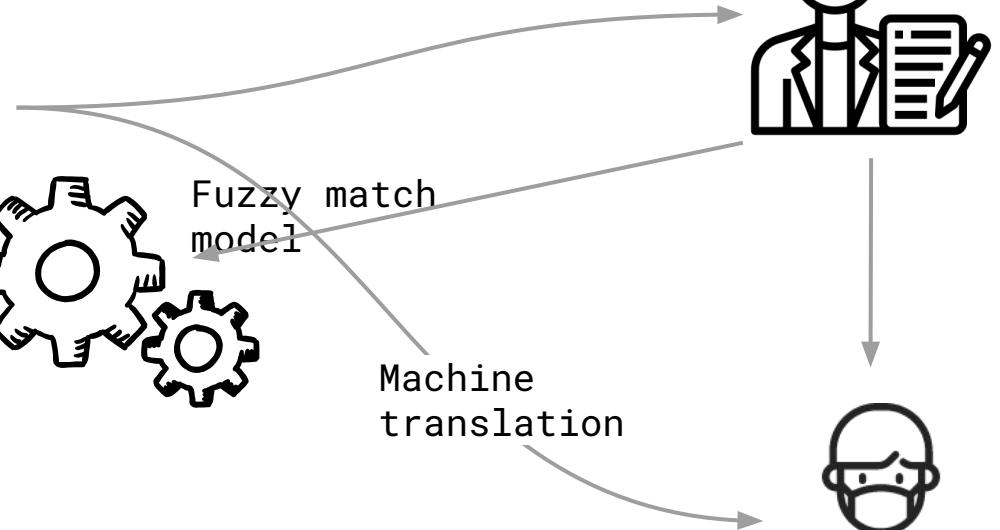
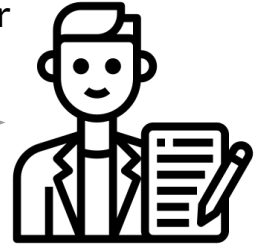
Hold the kidney medicine until you have a chance to speak with your kidney doctor.



Fuzzy match model

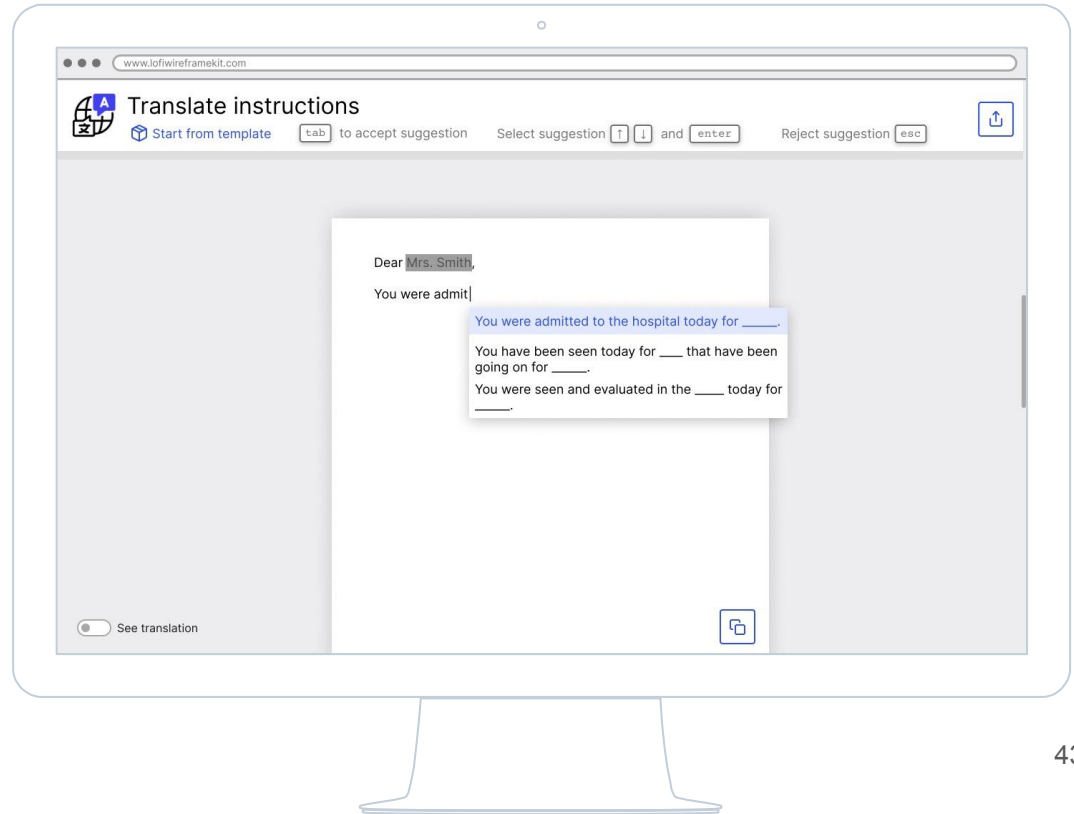
Machine translation

Professional translator



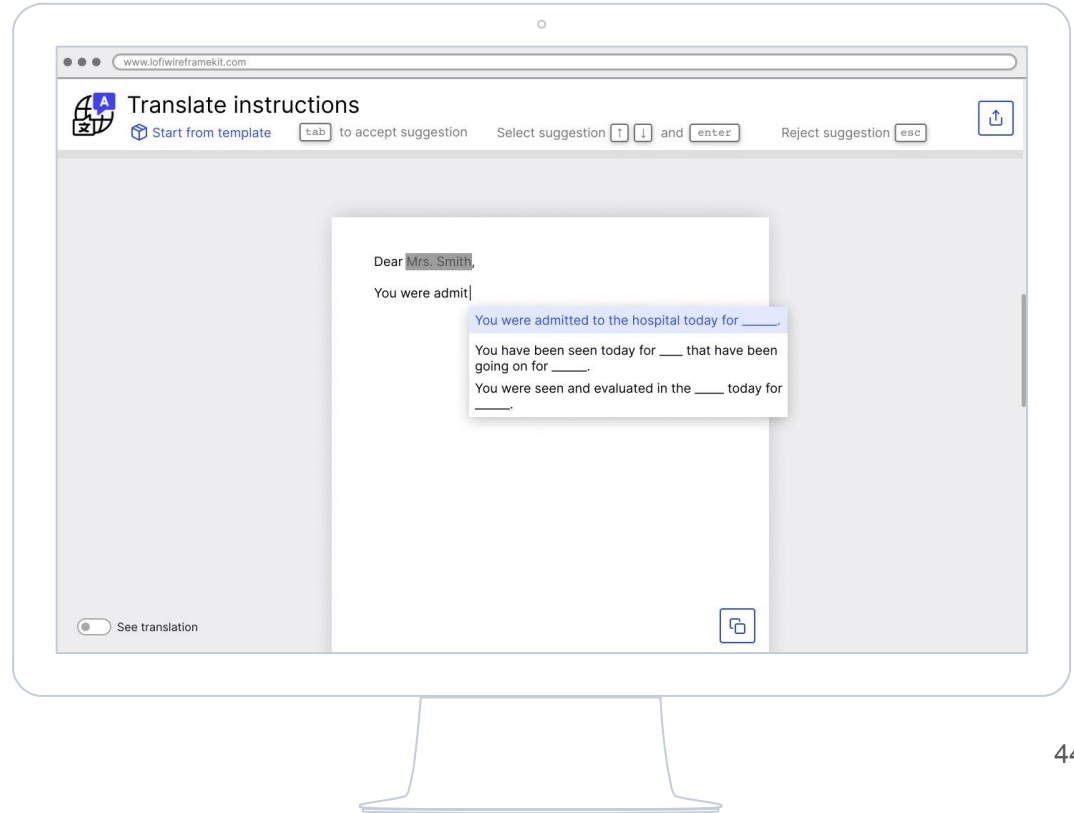
Design considerations

- Potential for errors
- Time constraints
- Acceptability to patients and clinicians



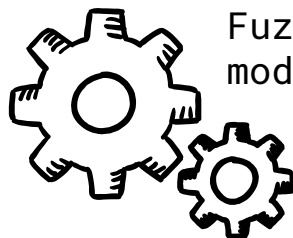
Design considerations

- **Potential for errors**
- Time constraints
- Acceptability to patients and clinicians

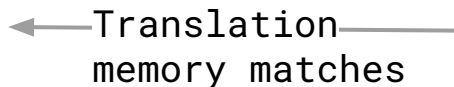


Using the translation memory

Hold the kidney medicine until you have a chance to speak with your kidney doctor.



Fuzzy match model



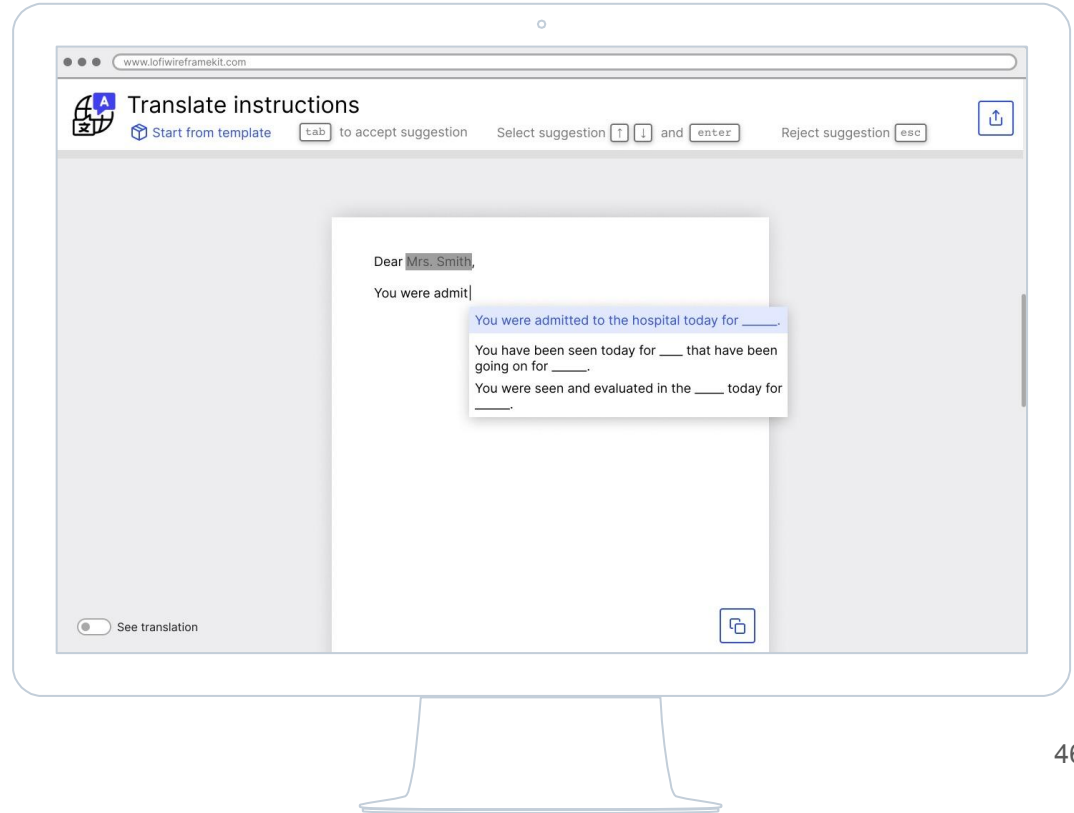
Keep taking kidney medicine until you talk to your kidney doctor

继续服用肾脏药物直到您有机会和您的肾脏医生交流



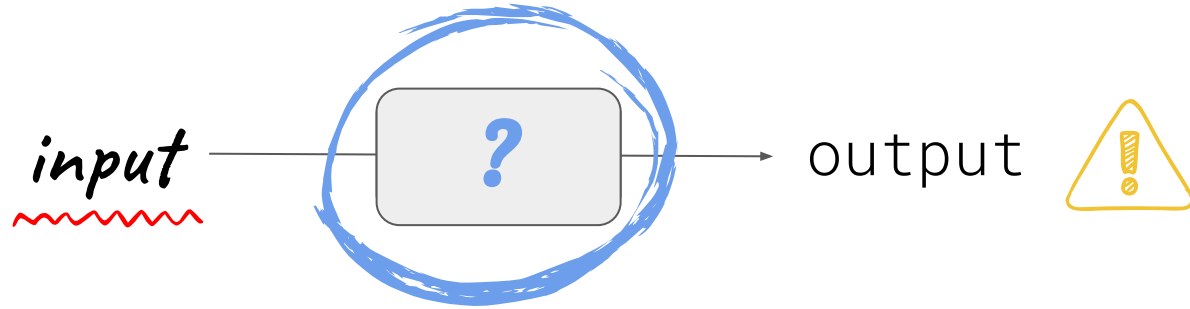
Design considerations

- Potential for errors
- Time constraints
- Acceptability to patients and clinicians

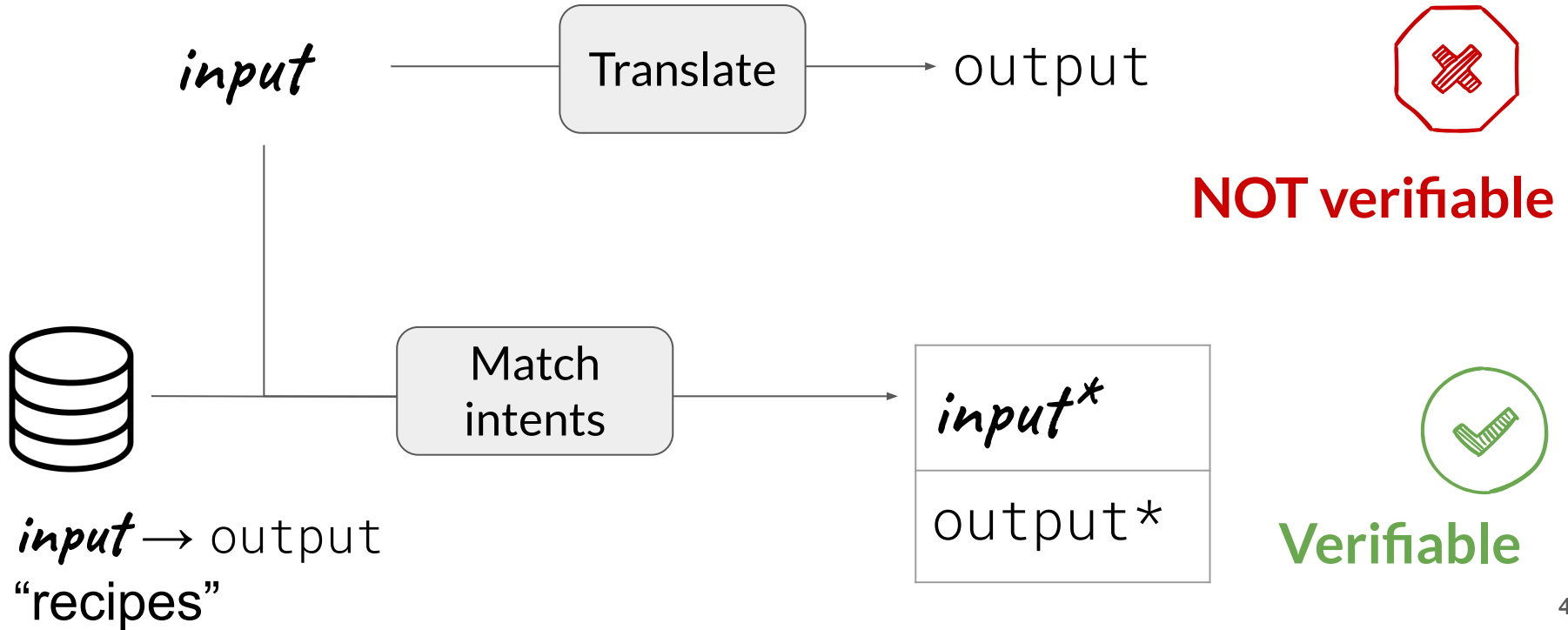


- Translation as a specification problem
- What do users need?
- Aside: Tools for professional translators
- Reframing the problem: from specification to retrieval
- **Takeaways**

Beyond bells and whistles



Reframing the problem



Thank you!

Elaine Khoong
Ana Milisavljevic
Sachin J. Shah
Wesley Hanwen Deng
Timnit Gebru
Margaret Mitchell
Daniel J. Liebling
Michal Lahav
Katherine Heller
Mark Díaz
Samy Bengio



Nikita Mehandru
Ph.D. Student,
School of Information



Kabir Goel
Undergraduate Student,
EECS



Niloufar Salehi
Assistant Professor,
School of Information &
EECS



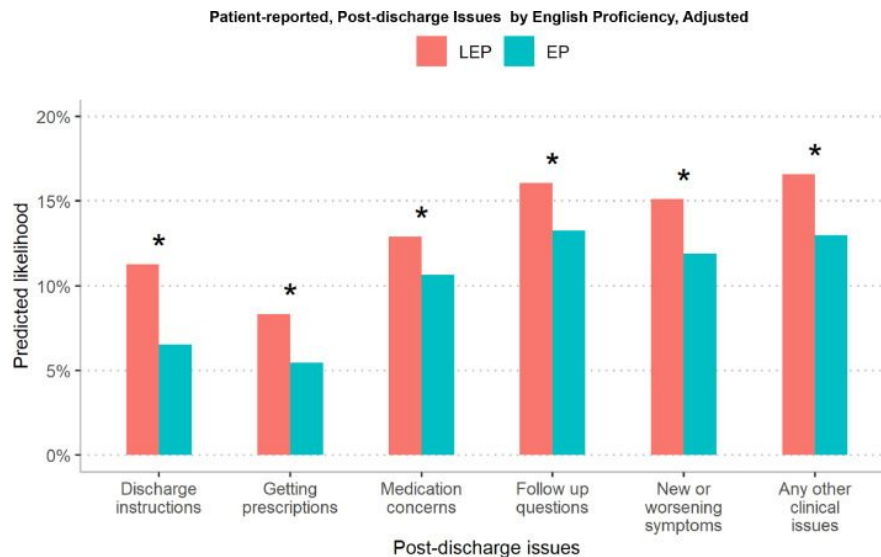
samantha_robertson@berkeley.edu



Lack of adequate translation support contributes to health disparities

“Patients with LEP were more likely to report a problem with all measured postdischarge issues”

[Malevanchik, et al., 2021]





>25 million

people in the United States have limited English proficiency