

# Responsible Gist MT Use in the Age of Neural MT

---

Marianna J. Martindale, iSchool PhD Candidate  
University of Maryland, College Park

Also: Computational Linguist, Center for Applied Machine Translation, USG

*OBLIGATORY DISCLAIMER: Opinions in this talk are my own and  
not necessarily those of any part of the U.S. Government*

# What Makes Neural MT (NMT) Different?

---

- Scores well on automated metrics & human evaluations
- Improves many types of errors (especially fluency)
- More languages & platforms than ever

But...

Sometimes fails catastrophically



# Humorous Catastrophic Failures

**Caren Marie Crabb**

Bark bark bark! Bark bark bark bark, bark.  
BARK!

| Good luck! God bless you. Good!

Automatically Translated

**Anne Marie Crabb**

BARK bark bark bark bark! barkbarkbark  
BARK!

| Good morning! God bless you!

Automatically Translated

**David & Daisy Finn**

bark bark bark. bark bark bark bark bark  
bark bark! bark...

| Good luck. It's good for you! Good luck...

Automatically Translated

Just now Like Reply More

**Heather Finner**

BARK!! Barkbarkbarkbark barkbark bark bark  
bark bark bark barkbarkbark!

| Blessed!! Bless you and bless you!

Automatically Translated

Facebook, 29 April 2020



# (Semi-)Humorous Catastrophic Failures

INTERNET NEWS JANUARY 18, 2020 / 12:27 PM / UPDATED 8 MONTHS AGO

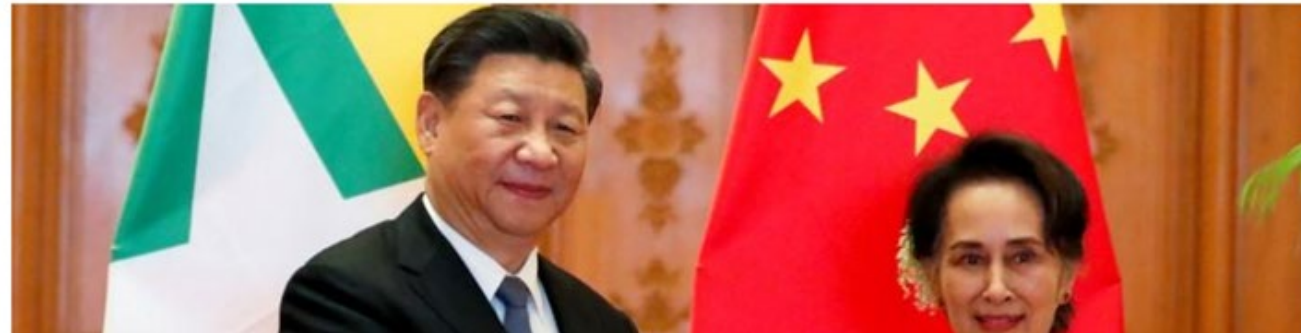


## Facebook says technical error caused vulgar translation of Chinese leader's name

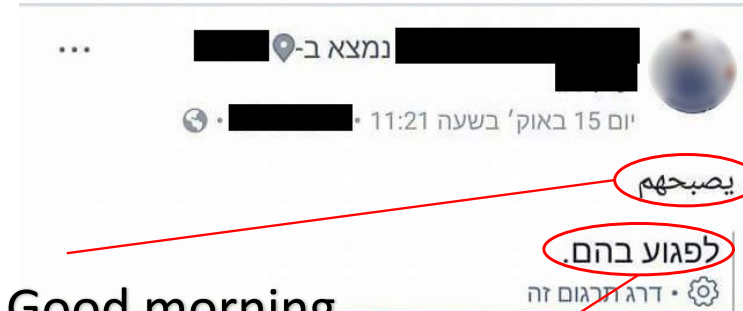
By Poppy McPherson

3 MIN READ

YANGON (Reuters) - Facebook Inc [FB.O](#) on Saturday blamed a technical error for Chinese leader Xi Jinping's name appearing as "Mr Shithole" in posts on its platform when translated into English from Burmese, apologizing for any offense caused.



# Dangerous Catastrophic Failures



يصبحهم = Good morning

לפגוע בהם ~ Attack them

<https://www.haaretz.com/israel-news/palestinian-arrested-over-mistranslated-good-morning-facebook-post-1.5459427>



# Dangerous Catastrophic Failures



**HAARETZ** | Israel News | All sections | Israel - BDS | Israel settlements | Italy - anti-Semitism | Flake - T

Home > Israel News

## Israel Arrests Palestinian Because Facebook Translated 'Good Morning' to 'Attack Them'

No Arabic-speaking police officer read the post before arresting the man, who works at a construction site in a West Bank settlement

Yotam Berger | Oct 22, 2017 1:36 PM

<https://www.haaretz.com/israel-news/palestinian-arrested-over-mistranslated-good-morning-facebook-post-1.5459427>



# When are (N)MT Errors Dangerous?

---

- Output is believable (in context)
- Lack of means and/or motivation to verify
- Use case involves MT informing action



# Believable Output

---

Believability = Fluency + Plausibility + Human Judgment

- Fluency: Does it “feel” like the target language?
  - Users more likely to trust fluent output (Martindale & Carpuat 2018)
  - NMT more likely to produce fluent but not adequate output (Martindale et al 2019)
- Plausibility: Does it make sense?
  - MT output is more believable when it is plausible (Work in progress)
- Human: People use heuristics to judge credibility of information<sup>1</sup>

<sup>1</sup>Rieh, S. Y. (2010). “Credibility and cognitive authority of information.” In M. Bates & M. N. Maack (Eds.), *Encyclopedia of Library and Information Sciences* (3rd ed., pp. 1337-1344).





# When are NMT Errors Dangerous?

---

- ✓• Output is believable (in context)
- Lack of means and/or motivation to verify
- Use case involves MT informing action

Gist MT?



# When are Gist MT Errors Dangerous?

---

Lack of means and/or motivation to verify?

Gist MT use characteristics

- High volume of foreign language text and/or tasks
- Impractical to translate everything or hire only bilinguals
  - Especially bilinguals with domain expertise
- Monolingual domain experts use MT to triage text or glean information
- Ideally: Bilinguals translate/evaluate documents/info monolinguals find
  - In practice: people may cut corners...



# When are Gist MT Errors Dangerous?

---

Use case involves MT informing action?

Gist MT use examples

- Journalist looking for relevant, local Tweets after an event
- Business analyst monitoring press for info about foreign competitors
- Investigator checking social media as part of background check



# Example: USCIS Refugee Vetting



## Appendix C: Translations

### Internet Translation Services

The most efficient approach to translate foreign language contents is to utilize one of the many free online language translation services provided by Google, Yahoo, Bing, and other search engines.

if needed. Use the following steps to translate using Google:

### In-Person Translation Services

Occasionally, officers will encounter foreign text written in a dialect or colloquial usage that does not necessarily translate easily using the available online tools mentioned above. Furthermore, there are currently no tools available to translate text written on images. Officers are responsible for determining

“Information collected from social media, by itself, will not be a basis to deny refugee resettlement”

Official statement, September 2019



# Example: USCIS Refugee Vetting

---

*“Information collected from social media, by itself, will not be a basis to deny refugee resettlement” --Official statement, September 2019*

However...

- Incorrect MT could tip scales of suspicion (in either direction)
- Social media is out of domain from MT training
- Often low-resource languages



# Is NMT for Gisting Worth the Risk?

---

- IMHO: Yes!

Good news:

- Truly misleading output is rare
- Faster to read, easier to understand
- Users like it

Just need to mitigate risk



# How can we mitigate the dangers?

---

## Dangers

- Output has errors
- Output is believable (in context)
- Lack of means and/or motivation to verify
- Use case involves MT informing action

## Mitigation goals

- ~~Error-free MT~~
- Encourage *appropriate* skepticism
- Make it easier to recognize potential errors
- Verify before acting



# How can we mitigate the dangers?

---

## Dangers

- Output has errors
- Output is believable (in context)
- Lack of means and/or motivation to verify
- Use case involves MT informing action

## Mitigation goals

- ~~Error-free MT~~
- Encourage skepticism
- Make it easier to recognize potential errors
- Verify before acting

**-Policy interventions**  
**-Technological Interventions**





# Mitigation Strategies

---

## Policy interventions

- Normative principles organizations with gist MT use cases should follow
- Changes to procedures and training

## Technological interventions

- Changes to the technology environment or the technology itself
- Requires additional research and development



# Policy Interventions

---

1. Independent, in-domain evaluation
2. Training for MT users
3. Workflows that require validation before action



# P1: Independent, In-Domain Evaluation

---

Principle: An organization should not deploy or encourage the use of MT without independent evaluation in the domain(s) and language pair(s) it is intended to be used on.

- If the intended use shifts/expands, additional testing should be conducted

Why? MT quality varies by language/domain

Independent – Not conducted by the MT company

Domain – Style and/or topic

Evaluation – Formal or informal

- Evaluators should know source language



## P2: Training for MT Users

---

Principle: Users should be trained to understand the technology well enough to expect variations in quality including dropped or hallucinated words and phrases.

Why? NMT is not intuitive! Hard to recognize what you don't expect.

Example hands-on exercises:

- Change context window, capitalization, punctuation, etc and observe output changes
- Compare output from high- and low- resource languages
- Try to get the system to hallucinate (e.g. fake Hawaiian)



# P3: Require Validation Before Action

---

Principle: Organizations with workflows that include critical decisions or actions informed by MT should require validation by someone who knows the source language before taking action.

Why? Establishing a consistent process deters corner-cutting.

- Even professional translation services rely on at least one level of quality control!

## Considerations

- Level of validation proportionate to impact of action/decision
- E.g., Self-validation through other resources *may* be sufficient for minimal-impact actions/decision



# Technological Interventions

---

1. Provide access to multiple MT outputs
2. Provide access to additional language resources
3. Build in “nudges” to help the user recognize quality issues



# T1: Multiple MT Outputs

---

What: Display outputs from two or more MT systems/models

LOE: Moderate

- Obtain licenses and/or build models
- Modify/create interface to display

Why? Users can observe differences to flag possible errors

Anecdote: Users actually prefer this anyway!



# T2: Additional Language Resources

---

What: Provide CAT-like tools to MT users

LOE: Low-Moderate

- Teach users features in existing services (e.g. Google Translate, Systran, Wiktionary, Linguee)
- Obtain access to resources (dictionaries/terminologies/TMs, etc)
- Integrate access alongside MT

Why?

- Individual word lookup can validate/clarify MT output
- Terminologies can resolve technical terms
- TM lookup can provide alternate contexts





# T3: Nudges

---

What: Automatically flag questionable output

- Quality estimation
- Diff on multiple outputs

LOE: High

- QE is an open research area

Why? Draw user's attention to problem areas



# Summary

Dangers	Mitigation goals	Recommended Interventions
<ul style="list-style-type: none"><li>• Output has errors</li></ul>	<ul style="list-style-type: none"><li>• <del>Error-free MT</del></li></ul>	<ul style="list-style-type: none"><li>• <i>(Continue improving)</i></li></ul>
<ul style="list-style-type: none"><li>• Output is believable (in context)</li></ul>	<ul style="list-style-type: none"><li>• Encourage <i>appropriate</i> skepticism</li></ul>	<ul style="list-style-type: none"><li>• P1 (Evaluation), P2 (Training), T3 (Nudges)</li></ul>
<ul style="list-style-type: none"><li>• Lack of means and/or motivation to verify</li></ul>	<ul style="list-style-type: none"><li>• Make it easier to recognize potential errors</li></ul>	<ul style="list-style-type: none"><li>• T1 (Multi-outputs), T2 (Lang resources), T3 (Nudges)</li></ul>
<ul style="list-style-type: none"><li>• Use case involves MT informing action</li></ul>	<ul style="list-style-type: none"><li>• Verify before acting</li></ul>	<ul style="list-style-type: none"><li>• P3 (Verify)</li></ul>

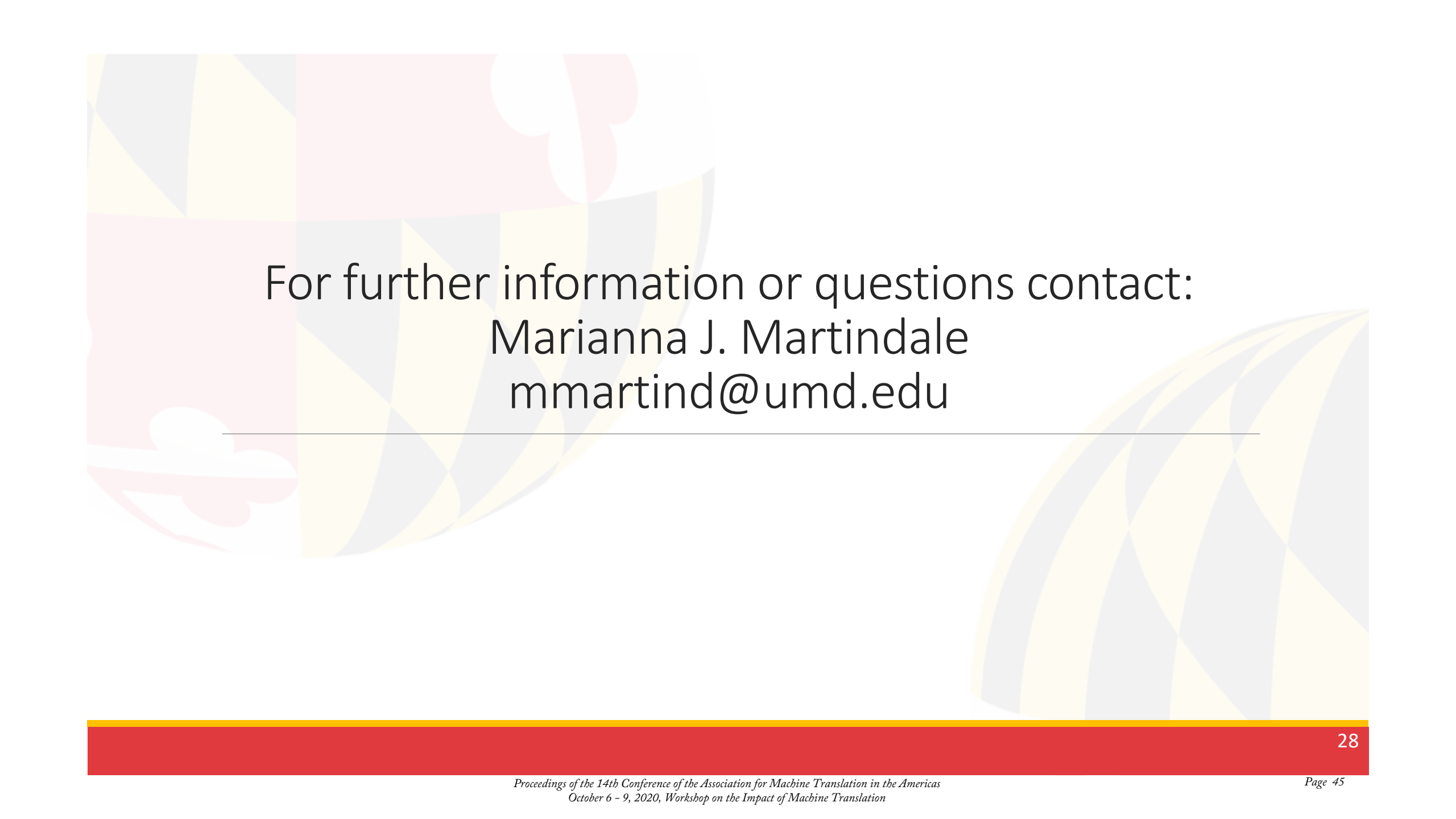


# Conclusion

---

- There can be risks to gist MT use
- Steps can be taken to mitigate them
- These are just examples
- Stakeholders should be looking at these mitigations and others
  - Organizational leadership
  - MT integrators
  - MT researchers
- See also: AI Ethics





For further information or questions contact:  
Marianna J. Martindale  
[mmartind@umd.edu](mailto:mmartind@umd.edu)

---