



# SUCCESSFUL TECH TRANSFER OF OF MT RESEARCH IN GOVERNMENT

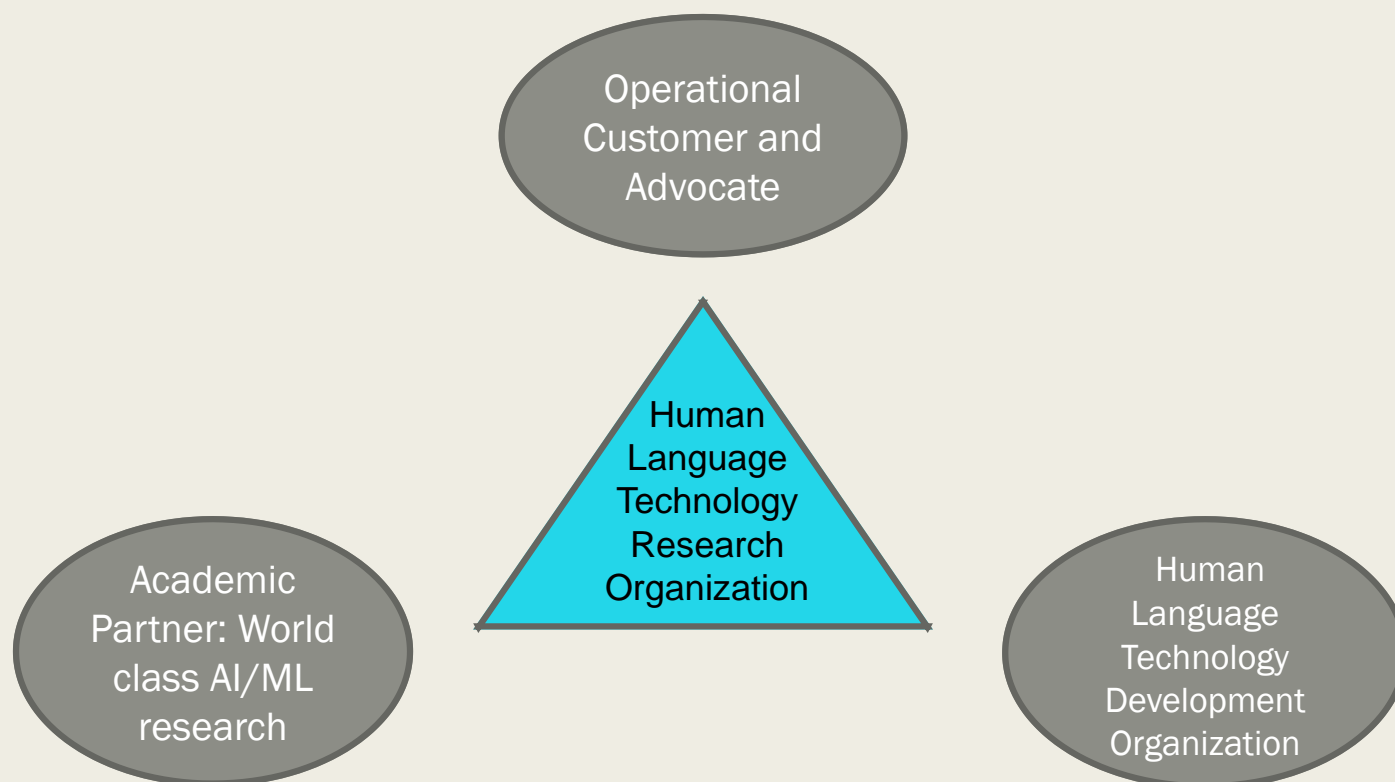
Dr. Kathy Baker  
US Dept. of Defense  
AMTA  
October 9, 2020

# Background

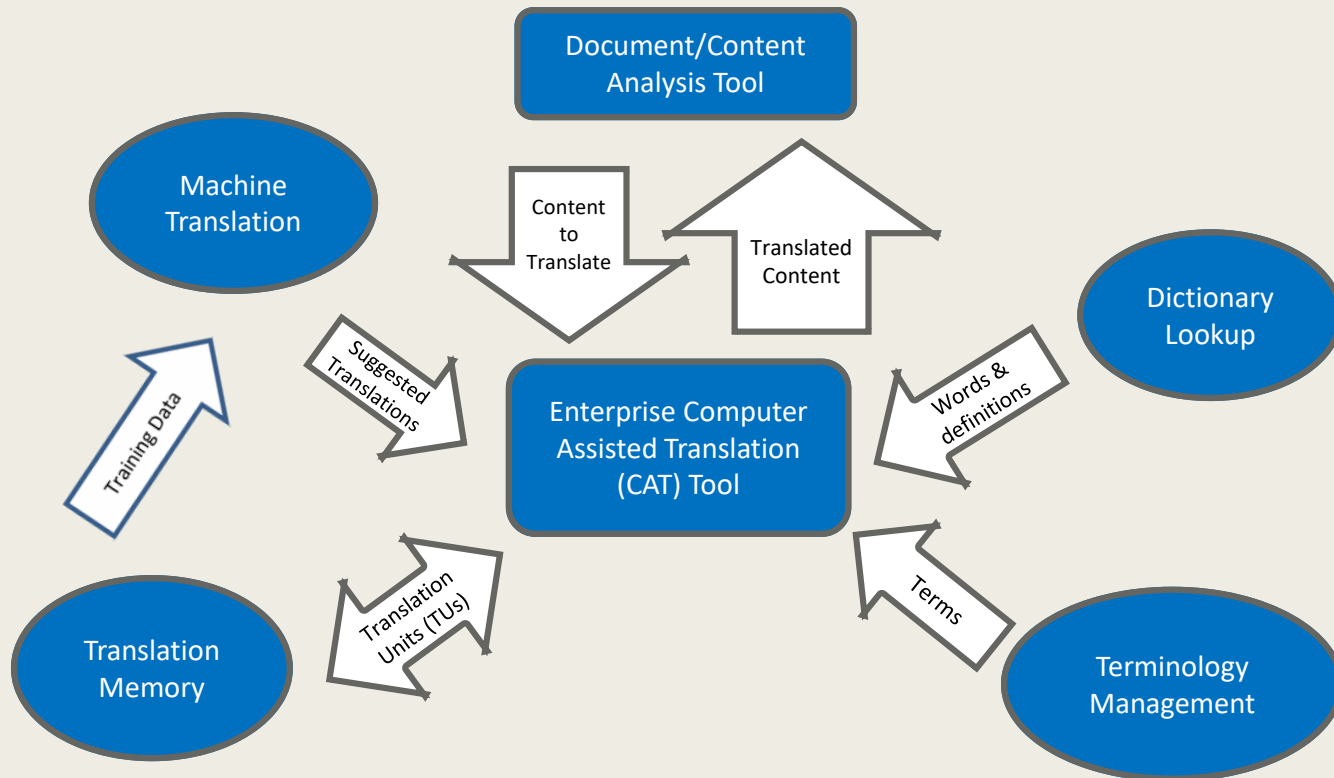
- *Vision* – DoD Agency recognizes Neural Machine Translation (NMT) as a force multiplier for analysis
- In-house HLT Research organization: Cutting edge Artificial Intelligence/Machine Learning (AI/ML)
- “Make vs. buy” – Agency not obligated to productize in-house research
- Key selling point is *access to and understanding of government data sets*

# What makes tech transfer successful?

- Partnerships, partnerships, partnerships!



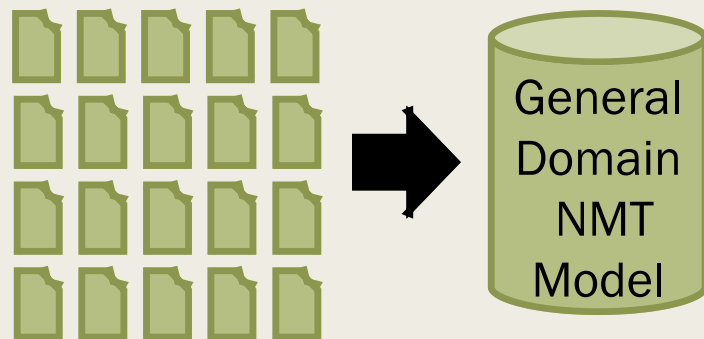
# Human Translation Technology Vision



# Academic Partnership

- Academic partnership with Johns Hopkins University
- Jump start research in neural MT with “dream team” of academic and industrial experts (Summer 2018 workshop)
- Adapt MT trained on general domain data for your use case with small amounts of highly technological or informal data

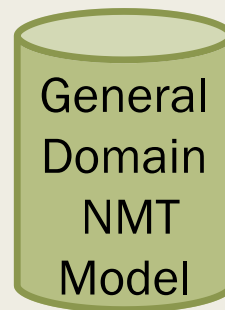
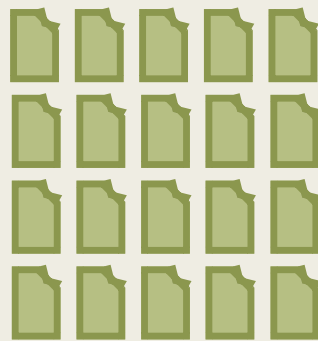
# General Domain Training



50M General-domain sentence pairs

General domain: News, parliamentary proceedings, movie subtitles, Wikipedia headlines, etc.

# General Domain Training



Russian patent example

Input: устройство для сбора воды с последующим её использованием для **стеклоомывателя**

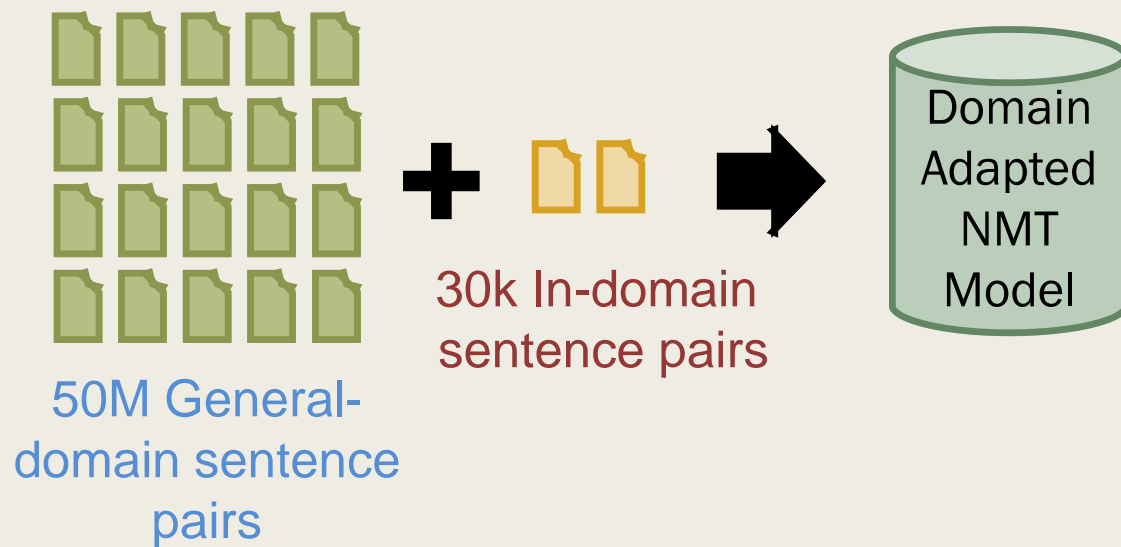
Human: device for collecting water for subsequent use in a **windscreen washer**

System: water collection device and subsequent use for **glazing**

Window  
glass

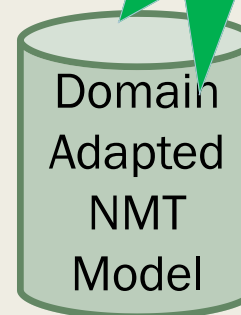
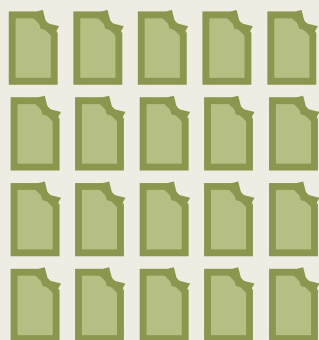
British for  
"windshield"

# Domain Adaptation





# Domain Adaptation



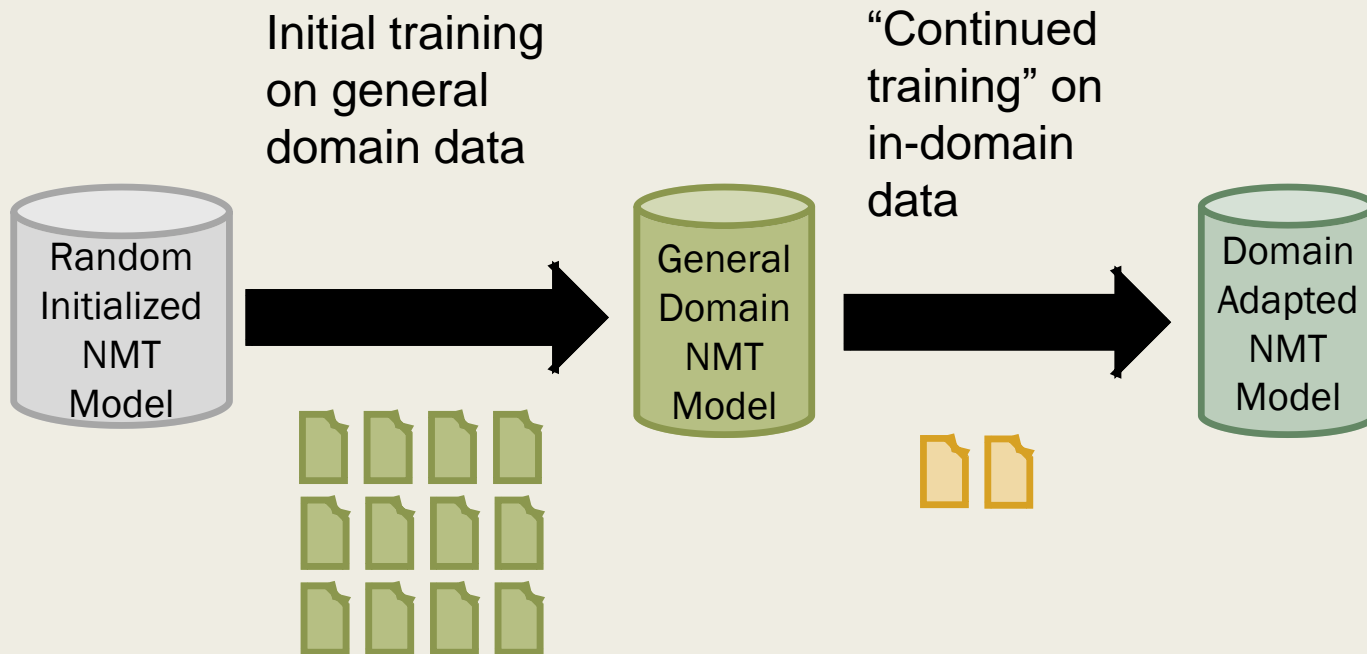
Input: устройство для сбора воды с последующим её использованием для **стеклоомывателя**

Human: device for collecting water for subsequent use in a **windscreen washer**

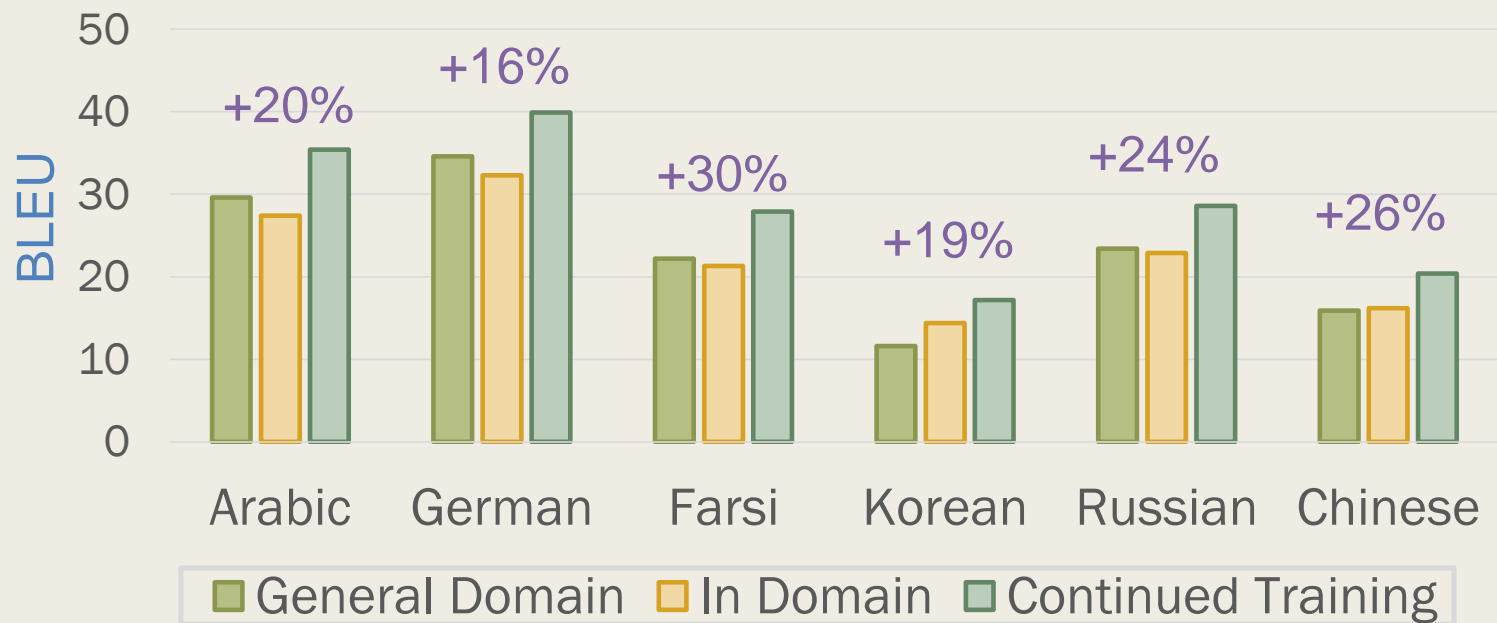
Output: water collecting device with subsequent use thereof for a **windscreen washer**

# Continued Training

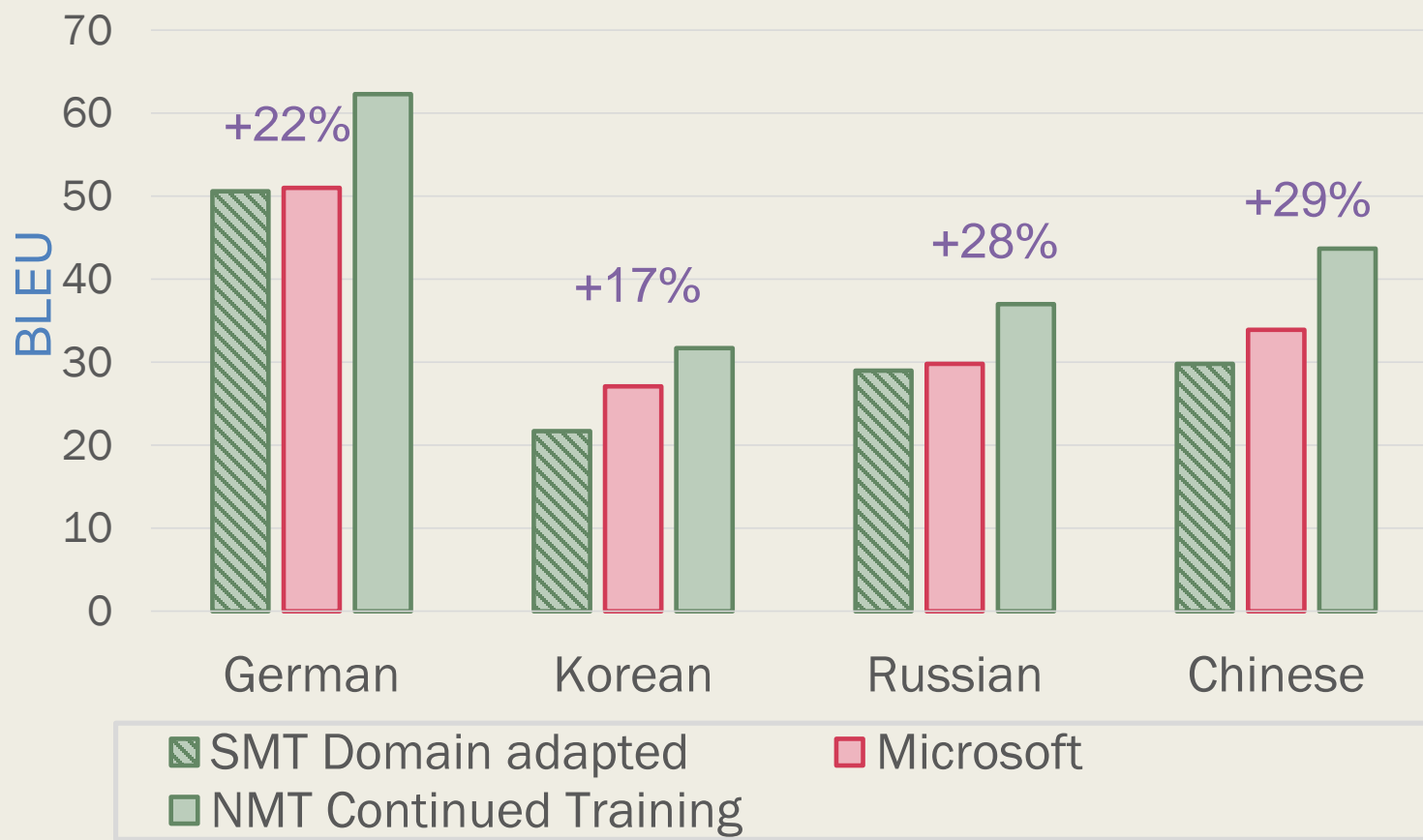
[= “Fine-tuning” in STT literature]



# TED Corpus – NMT Continued Training



# Patents - NMT vs Statistical MT, Out of box MT



# Sample Translations

## TED Talk

- **Source:** 我遇到了霍金教授 他说他的梦想是空间旅行
- **Reference:** And I met Professor Hawking, and he said his dream was to travel into space.
- **Agency Baseline:** I ran into Professor Hodgen he saying that his dream was the spatial travel
- **DA-NMT:** I met Professor Hawking who said that his dream was space travel.

## Patent

- **Source:** 一种基于纳米滤膜的废有机溶剂处理工艺及系统
- **Reference:** WASTE ORGANIC SOLVENT TREATMENT PROCESS AND SYSTEM BASED ON NANO FILTRATION MEMBRANE
- **Agency Baseline:** One kind based on nanometer filter diaphragm's waste organic solvent processing craft and system
- **DA-NMT:** Waste organic solvent treatment process and system based on nanometer filter film

# Operational Partnership

- Customer who understands benefits of research early on contributes
  - *Advocacy*
  - *Funding*
  - *Personnel with language expertise*
  - *Operational Data*

# Development Partnership

- HLT Development Organization designed as a “sister org” with common goals for success
- Previous Research successes built up model deployment infrastructure there
  - *Own and provide necessary hardware*
  - *Own the enterprise wrapper for machine translation; no separate User Interface needed*

# Open Source Software

- Several well-designed platforms for building neural models
- Cost savings on initial push to Operations (though beware the maintenance tail)
- **Allows control of training data and model building**



# Current Status

- Six languages deployed in beta status
- Very well received
- What's left?
  - *Scaling! Throughput, architecture*
  - *More languages*
  - *Tradcrafft: Feedback loop from translation at workstation to model retraining*
  - *Modernizing analyst translation interface \*in tandem\* with deploying to current portal*