DARPA
DARPATech
2002 Symposium
Transforming
Fantasy

Human Language Technology TIDES, EARS, Babylon

Charles L. Wayne

Transforming Fantasy

DARPA

Human Language Technology

Distribution Statement A. Approved for public release; distribution is unlimited.

TIDES

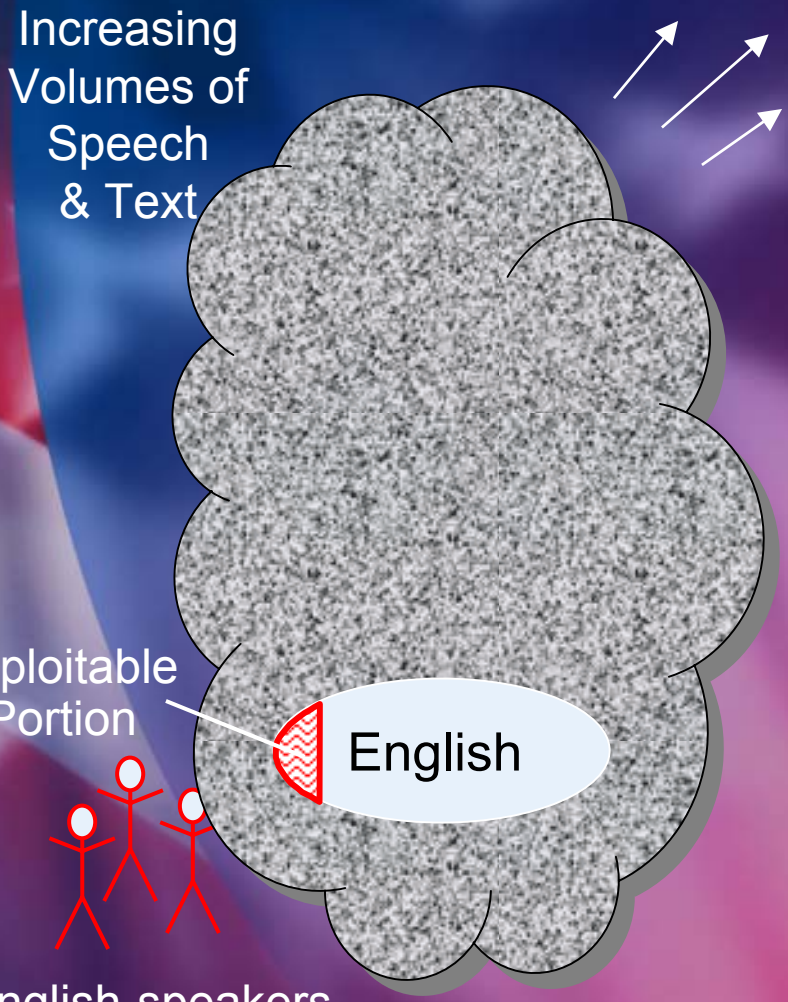
Translingual Information Detection, Extraction and Summarization

Transforming Fantasy

DARPA

Distribution Statement A. Approved for public release; distribution is unlimited.

Before TIDES

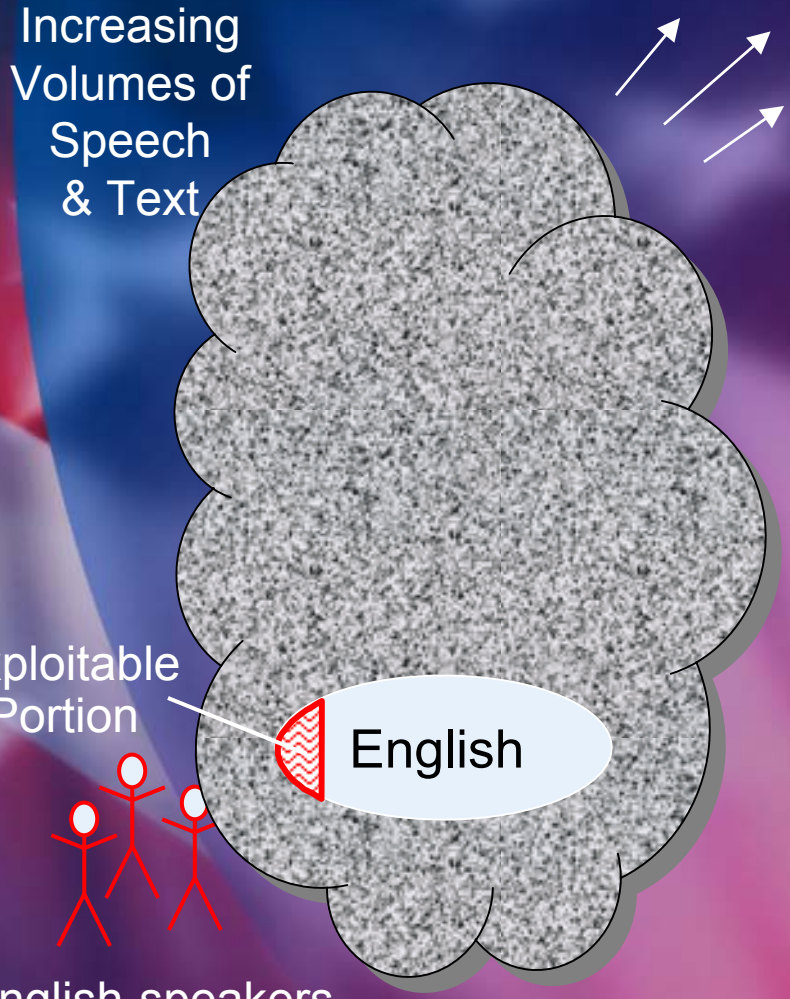


Transforming Fantasy

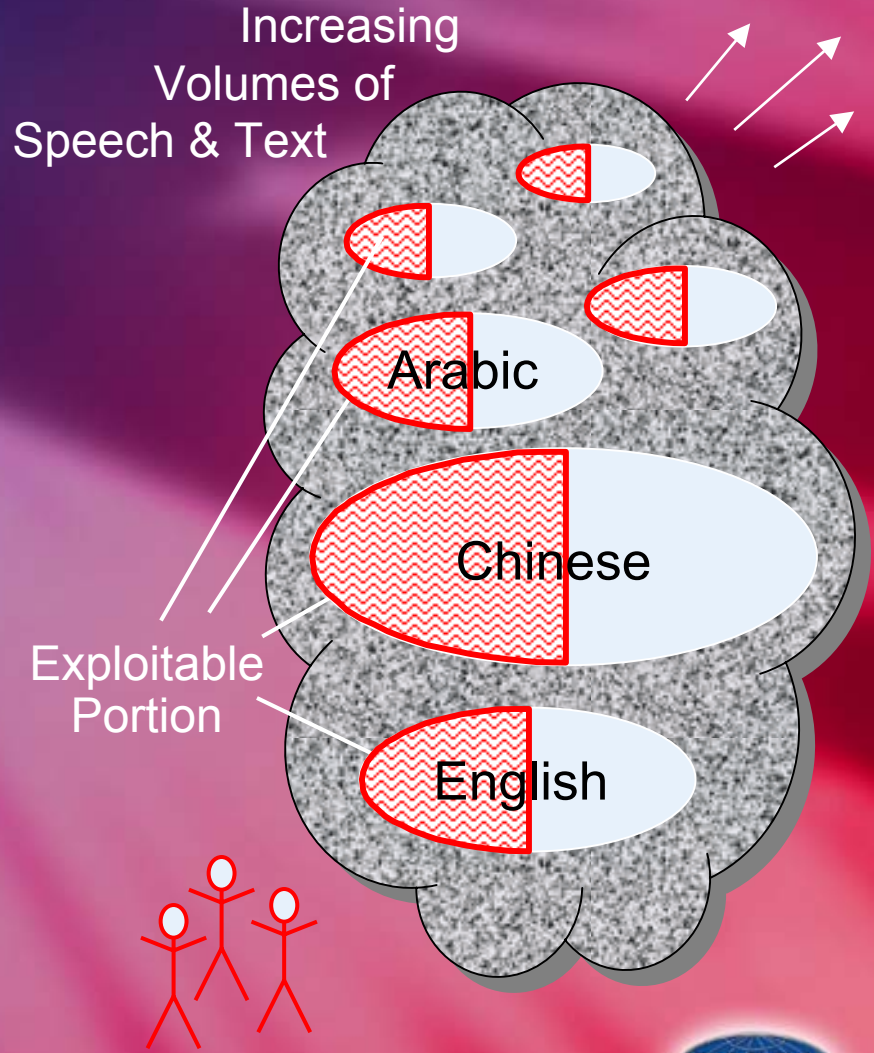
DARPA

Distribution Statement A. Approved for public release; distribution is unlimited.

Before TIDES



After TIDES



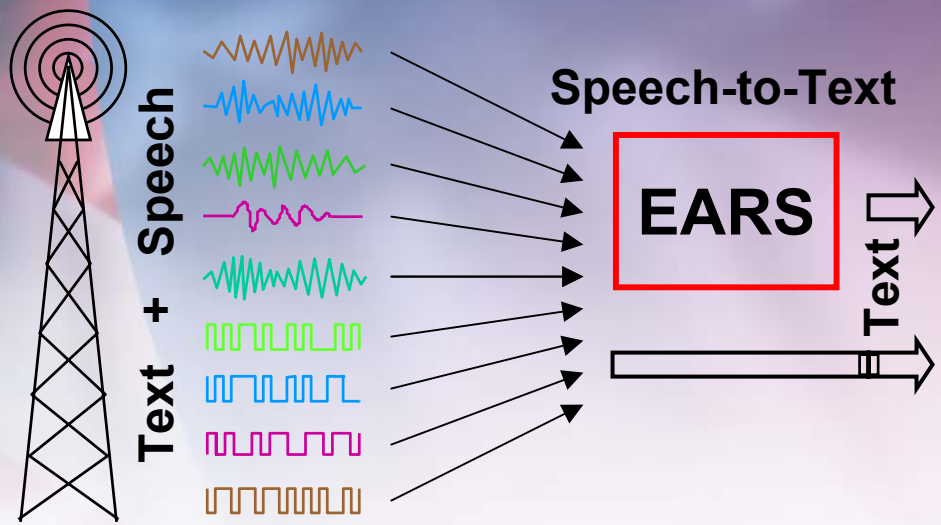
Transforming Fantasy

DARPA

Distribution Statement A. Approved for public release; distribution is unlimited.

Goal

Multiple Media
Multiple Sources
Multiple Languages

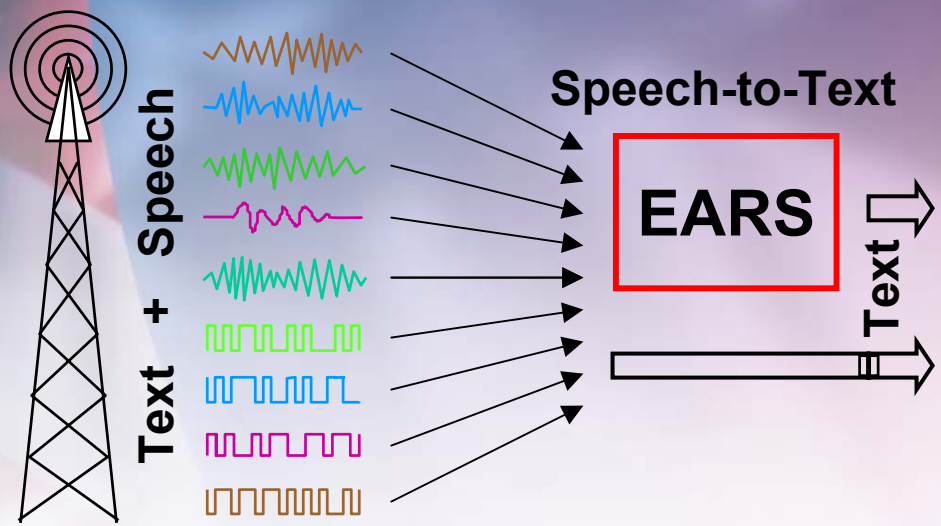


Streaming, Changing
Huge Volumes

Distribution Statement A. Approved for public release; distribution is unlimited.

Goal

Multiple Media
Multiple Sources
Multiple Languages



Streaming, Changing
Huge Volumes

Policy Makers
& Warfighters

English-speaking
operator/
analyst

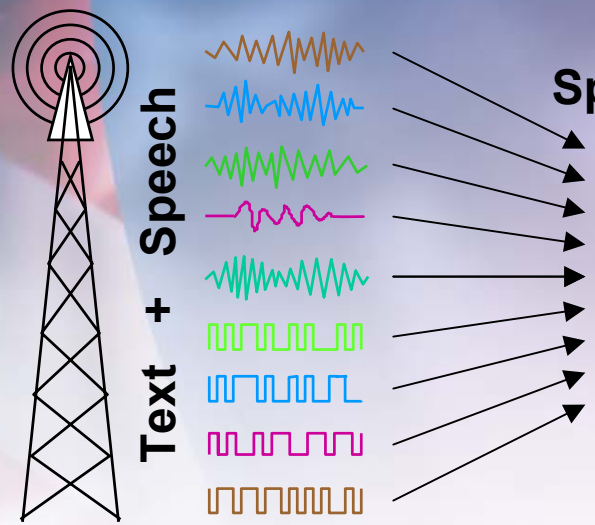
Distribution Statement A. Approved for public release; distribution is unlimited.

Goal

Multiple Media
Multiple Sources
Multiple Languages

Policy Makers
& Warfighters

English-speaking
operator/
analyst



Speech-to-Text

EARS

Text

TIDES

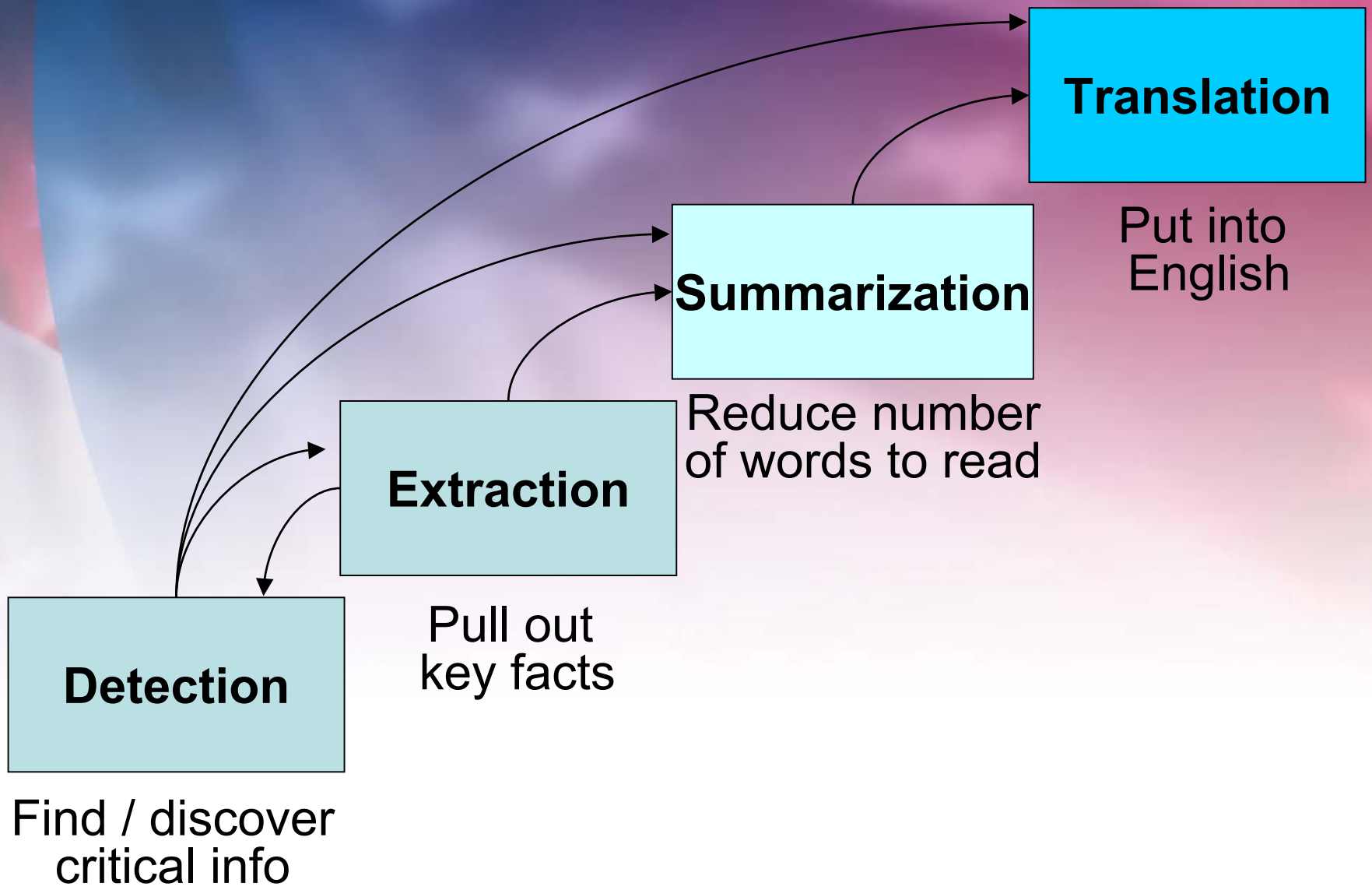
Critical
Information
(in English)

Streaming, Changing
Huge Volumes

English Speakers Able to Find & Interpret Mission-Critical Information

Distribution Statement A. Approved for public release; distribution is unlimited.

Component Technologies



Distribution Statement A. Approved for public release; distribution is unlimited.

Research Progress

- Detection
  - Cross-language information retrieval
  - Topic tracking
  - New event detection
- Extraction
  - Entities
  - Relationships
- Summarization
  - Document compression
  - Cluster description
- Translation
  - Novel approaches evaluation

Transforming Fantasy

DARPA

Distribution Statement A. Approved for public release; distribution is unlimited.

Languages

English
Chinese
Arabic

Principal focus

Ultimate goal:
Cheap/rapid porting to new languages

Transforming Fantasy

DARPA

Distribution Statement A. Approved for public release; distribution is unlimited.

Technology Integration, Experimentation, Transition

- Text and Audio Processors (TAP systems)
- Experiments on real-world problems with real data and real users
- Transition of successful technology

Transforming Fantasy

DARPA

Distribution Statement A. Approved for public release; distribution is unlimited.

Systems Developed

- Open Audio Source Information System (OASIS)
- On-Line Text and Audio Processor (OnTAP)
- MITRE Text and Audio Processor (MiTAP)

Transforming Fantasy

DARPA

Distribution Statement A. Approved for public release; distribution is unlimited.

Integrated Feasibility Experiments

- IFE – Bio
- IFE – Arabic
- IFE – Translingual

Transforming Fantasy

DARPA

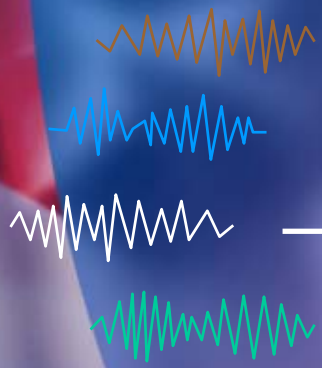
Distribution Statement A. Approved for public release; distribution is unlimited.

EARS to the World

Effective, Affordable, Reusable Speech-to-Text

Distribution Statement A. Approved for public release; distribution is unlimited.

Powerful speech-to-text technology



Human-human speech

Rich Transcription

Multiple Applications

WORDS + METADATA

Rich transcript

Transforming Fantasy

DARPA

Distribution Statement A. Approved for public release; distribution is unlimited.

Payoff

- Enormous force multiplier –
  - Machines able to
    - –Detect much better
    - –Extract, summarize, translate
  - Humans able to
    - –Read & understand easily, quickly

Transforming Fantasy

DARPA

Distribution Statement A. Approved for public release; distribution is unlimited.

BABYLON

Two-way speech-to-speech translation
for the combat environment

Transforming Fantasy

DARPA

Distribution Statement A. Approved for public release; distribution is unlimited.

Speech-to-Speech Translation

One-Way

Phrase recognition, playout
Used in Kosovo, Afghanistan
("Phraselator")

One-plus-One-Way

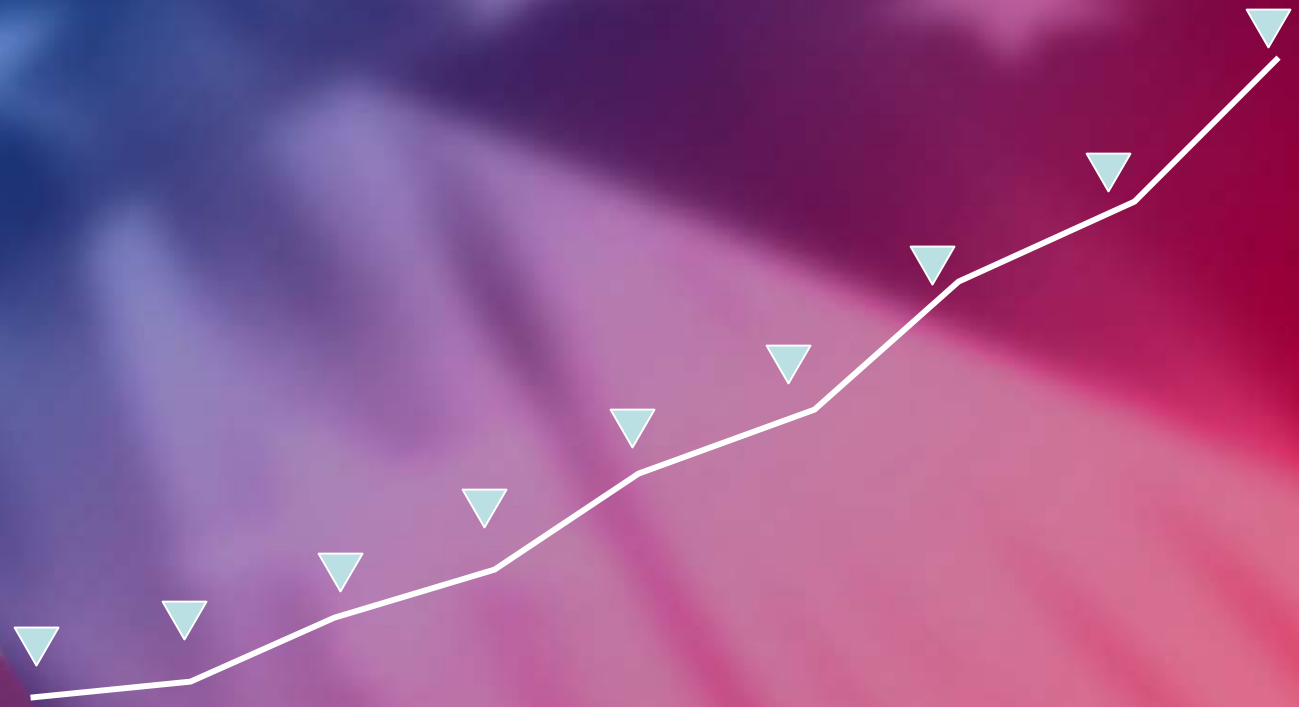
Enable English-speakers to understand others

Two-Way

Greater freedom of expression

Distribution Statement A. Approved for public release; distribution is unlimited.

Objective Evaluations Drive Progress



Transforming Fantasy

DARPA

Distribution Statement A. Approved for public release; distribution is unlimited.

Key Evaluations

- Rich Transcription
- Information Retrieval
- Topic Detection & Tracking
- Automatic Content Extraction
- Summarization
- Machine Translation

Transforming Fantasy

DARPA

Distribution Statement A. Approved for public release; distribution is unlimited.

Summary

- Human Language Technology
  - Key enabler
  - Hear, read, understand
  - Total Information Awareness
- Key Programs
  - TIDES
  - EARS
  - Babylon
- Evaluation

Transforming Fantasy

DARPA

Distribution Statement A. Approved for public release; distribution is unlimited.

DARPA
DARPATech
2002 Symposium
Transforming
Fantasy