

IMPLEMENTING ARTIFICIAL INTELLIGENCE SOLUTIONS FOR COMMON TEST DEVELOPMENT CHALLENGES

THE TWENTIETH ANNUAL MARC CONFERENCE:

**KIMBERLY SWYGERT
IAN MICIR
NBME
NOVEMBER, 2022**

WHAT NBME DOES



- ∅ Protects the health of the public through state-of-the-art assessment
- ∅ Flagship product: USMLE series (Steps 1-3) for initial physician licensure
- ∅ Massive volume – extensive item bank review and maintenance
- ∅ Test Development: ~100 staff, extensive item development and review processes; support vigilant test security
- ∅ State of the art is a state of mind



AI IN ASSESSMENT

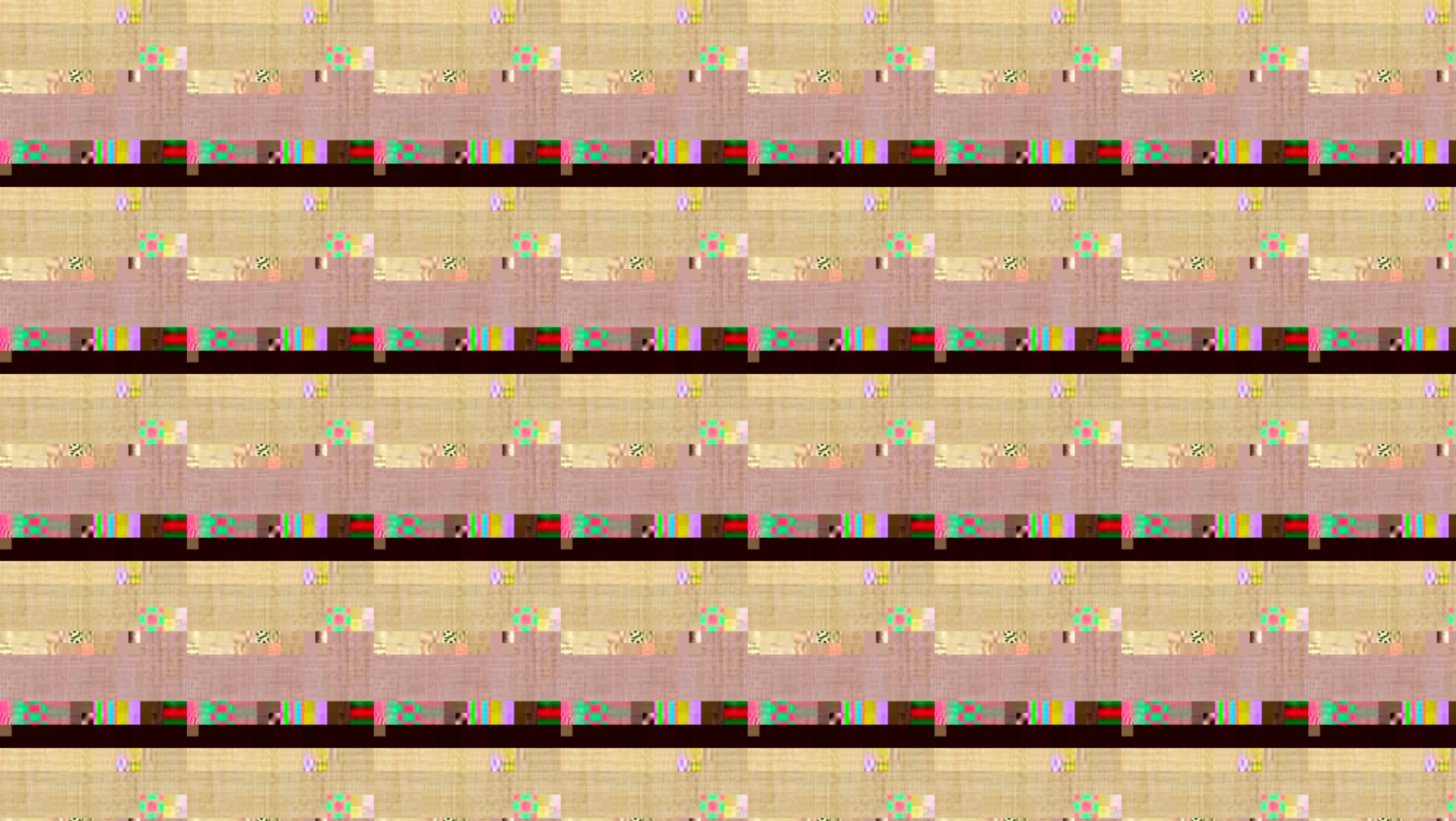


- ∅ AI within educational fields, including medical education, is moving forward at a rapid pace, where AI tools are being deployed within the context of learning and tutoring
- ∅ AI within assessment, including licensure assessment, is being implemented more slowly than the educational or field of practice uses of AI
- ∅ Recently, there have been promising indicators:
 - *Principles of AI Use In Testing* (Association of Test Publishers, 2021)
 - *Computational Psychometrics* (von Davier, Mislevy, & Hao, 2022)
 - Pandemic response necessitating AI tools, such as AI proctoring

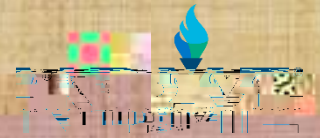
AI IN ASSESSMENT



- ∅ However, the challenges are substantial, with technology and data not always the biggest challenges
- ∅ Test developers abide by the *Standards*, while the use of AI introduces a new set of ethical standards that overlap but are not identical
- ∅ Regulation of data use for AI may be in addition to regulations for responsible governance and use of data within testing organizations



AI AT NBME



∅ In alignment with this, we focused on:

- Coordination with AI research efforts and other business uses of AI
- Coordination with internal data strategy, privacy, and governance efforts
- Scanning the landscape for AI use in medical education and practice
- Developing of an AI Center of Excellence
- Planning for organizational strategy that would guide future efforts

∅ We also realized that crucial ethical concerns may emerge, even with AI development on a small scale where examinees were not directly interacting with the AI

ETHICAL CAUTIONS AND OUR AI PRINCIPLES



Most people who bother with the matter at all would admit that the English language is in a bad way, but it is generally assumed that we cannot by conscious action do anything about it. Our civilization is decadent and our language ^{is} so the argument runs ^{must} inevitably share in the general collapse. It follows that any struggle against the abuse of language is a sentimental archaism, like preferring candles to electric light or hansom cabs to aeroplanes. Underneath this lies the half-conscious belief that language is a natural growth and not an instrument which we shape for our own purposes.

Now, it is clear that the

ETHICAL CAUTIONS AND OUR AI PRINCIPLES



∅

∅

it is cool!

∅

In order: *Should we do it? Can do we it? How should we do it?*

∅

Moral responsibility and respect for human dignity

∅

∅

Our goal: accuracy, efficiency, and employee satisfaction

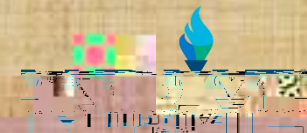


OUR AI APPLICATIONS – YOGI



- Ø .NET application written in C#
- Ø Compares blurbs found on the internet to our item banks
- Ø Automates a good deal of pre-editing via Regex patterns
- Ø Uses TF-IDF vectors to calculate cosine similarity (stems & answers)
- Ø Metrics: Previous method vs. Yogi
- Ø Slides from 2019 presentation at Conference on Test Security (COTS) available upon request

OUR AI APPLICATIONS – YOGI



Previously presented at COTS, 2019

OUR AI APPLICATIONS – YOGI



The screenshot displays the Maeve AI application interface. The window title is "Maeve". The main content area shows a complex interface with a table of items, a "Confidence" section with radio buttons for "Low", "Medium", and "High", and a "Notes" section. A specific note reads: "Ian Micir: The answer is exact match. I could use a second opinion on the item stem." At the bottom right, there are "Reject" and "Approve" buttons.

MatchItem	TargetItem	STEM	OPTIONS	ANSWER	STATUS
MBX7354	03-WM46	0.1588	0	0	Rejected

Previously presented at COTS, 2019

OUR AI APPLICATIONS – SMOKEY



- ∅ .NET application written in C#
- ∅ Extension of Yogi NLP engine with ML element added
- ∅ Parameters: stem & answer similarity, n content codes, enemy status
- ∅ Supervised learning: existing enemy status = truth variable
- ∅ Metrics included editor satisfaction and reduction of hours needed to complete the tasks in addition to accuracy
- ∅ Published in the *Journal of Applied Testing in Technology*, 2002

∅

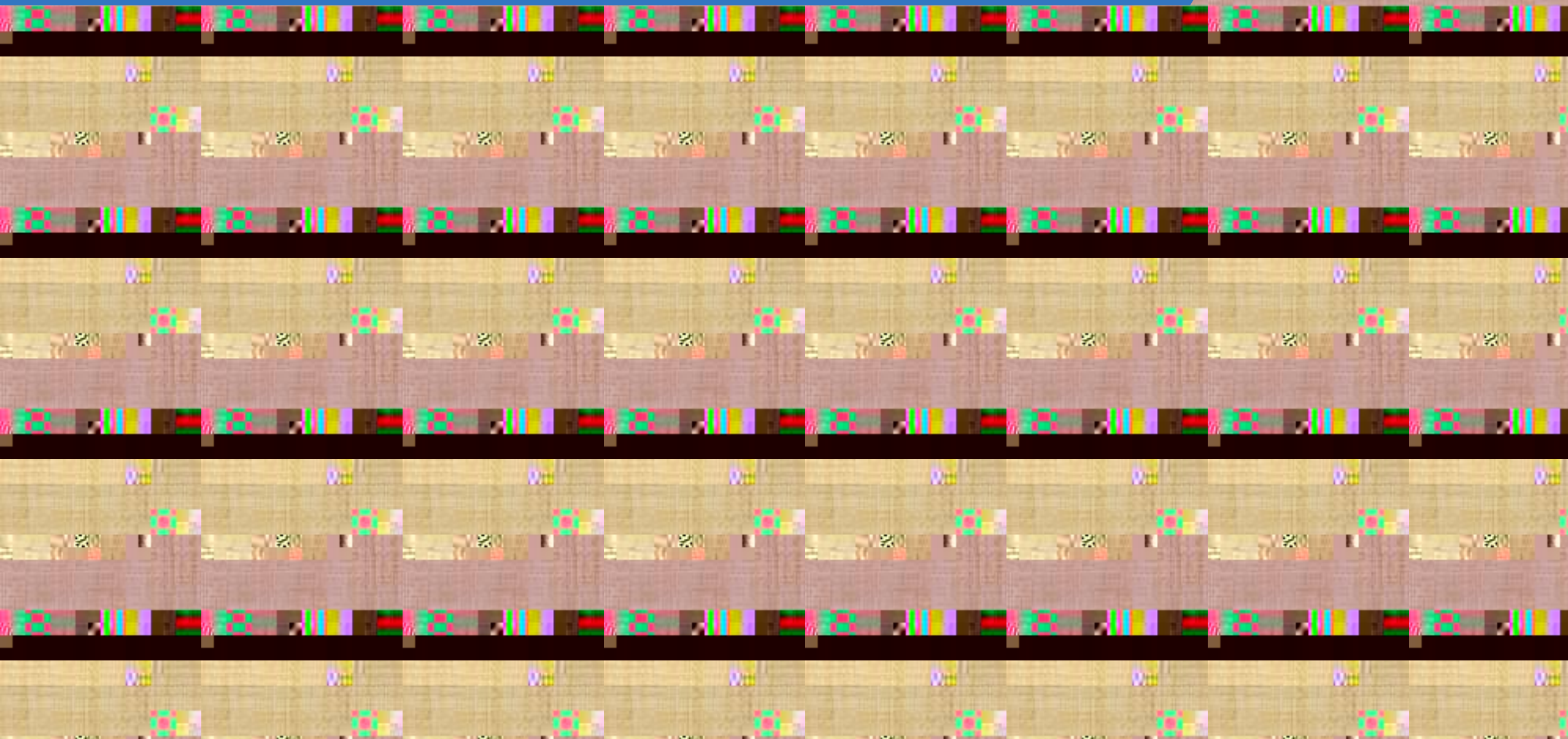
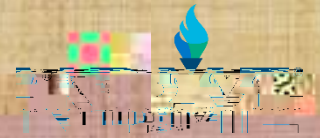
OUR AI APPLICATIONS – SMOKEY



1	100	0	0	0	100
2	73	27	0	0	100
3	84	16	0	0	100
4	91	9	0	0	100
5	97	3	0	0	100
6	79	21	0	0	100
7	66	0	23	11	100
TOTAL	590	76	23	11	700

1	7	3	14	1	25
2	7	6	10	0	23
3	12	22	12	0	46
4	33	15	4	0	52
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	0	0
TOTAL	59	46	40	1	146

OUR AI APPLICATIONS



Kimberly Swygert, kswygert@nbme.org
Ian Micir, imicir@nbme.org

THANK YOU!