

Towards Comprehensible Modeling

Michael Gleicher

Department of Computer Sciences

University of Wisconsin Madison



Acknowledgements

This work is a collaboration with a great set of people!

Students:

Danielle Albers, Eric Alexander, Michael Correll,
Adrian Mayorga, Alper Sarikaya, Erin Winter, ...

Collaborators:

Steve Franconeri, Jonathan Hope, Robin Valenza,
Mike Witmore, Remco Chang, Bilge Mutlu, ...

This research is funded, in part, by the National Science Foundation, NIH, and the Andrew Mellon Foundation.

Towards Comprehensible Modeling

Michael Gleicher

Department of Computer Sciences

University of Wisconsin Madison



Human Centric Issues in Data Analytics

~~Towards Comprehensible~~

~~Modeling~~

Michael Gleicher

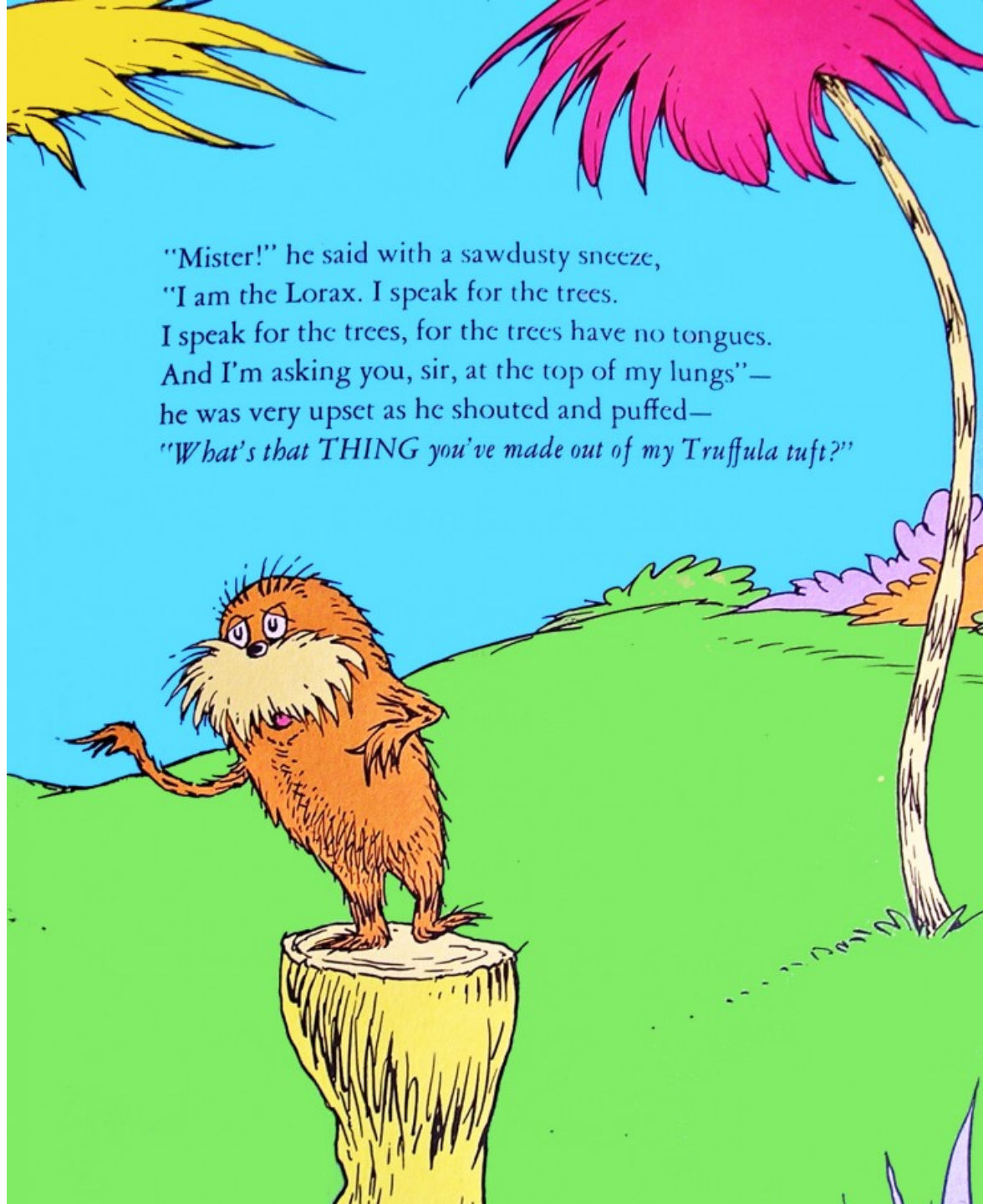
Department of Computer Sciences

University of Wisconsin Madison



I am the Vis Guy
I speak for the Users ...

"Mister!" he said with a sawdusty sneeze,
"I am the Lorax. I speak for the trees.
I speak for the trees, for the trees have no tongues.
And I'm asking you, sir, at the top of my lungs"—
he was very upset as he shouted and puffed—
"What's that THING you've made out of my Truffula tuft?"



Mike's Soapbox

Human-Centric Issues Matter

Comprehensible models are a way at getting at (many of) them

Many different ways to get at
Comprehensibility in Modeling

Why?

Do we do “this”?

(data and data analysis)

Why?

Do we do “this”?

To help someone do something

(but how do you think about that?)

Who, Why, Where...

Who?

Stakeholders

Developers

Data Scientists

Domain Experts

Audience

Subjects

Why?

Reason for Wanting

Improve Performance

Build Theory

Extend/Characterize

Build Trust

Actionability

Convince

Where?

Phase of the **Process**

Identify Problem

Gather Data

Clean Data

Abstract

Design

Build

Evaluate

Interpret

Act

Disseminate

So What?

Consider everybody

Consider the range of Goals

Consider helping in different phases of the process

Consider helping in one phase to effect another

There's such a wide range of situations that there is unlikely to be one "best" approach

Why Modeling? (as a place to focus)

It's part of most data applications

If you have lots of data, you (usually) need to reduce it to a human usable scale

It's a useful tool generally

It's increasingly hard to comprehend

Ways to help comprehensibility

Look Inside the Model

- Better examination tools

- Models designed to be examined

Look Outside the Model

- Examine the outcomes

- Task-centric tools

Is this a useful distinction?

Some Examples

Use Simple Models for Insight

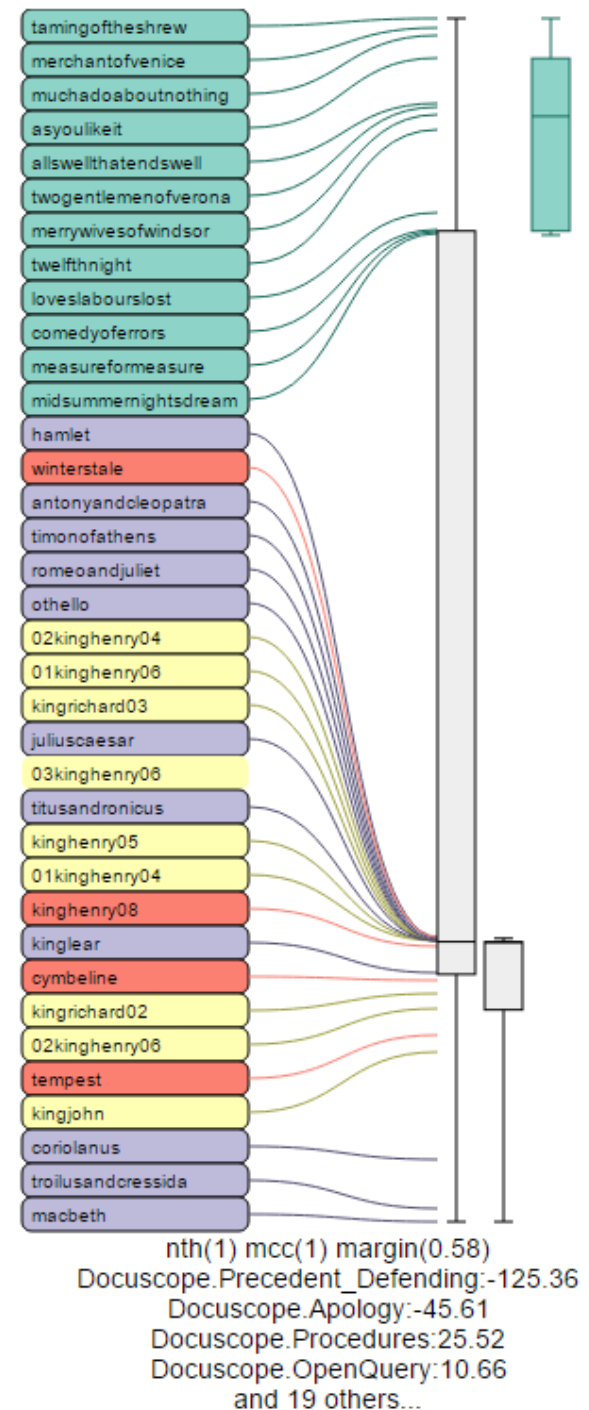
Build classifiers for known groups

Look “inside” the models to see how they work

Create models specifically for comprehensibility

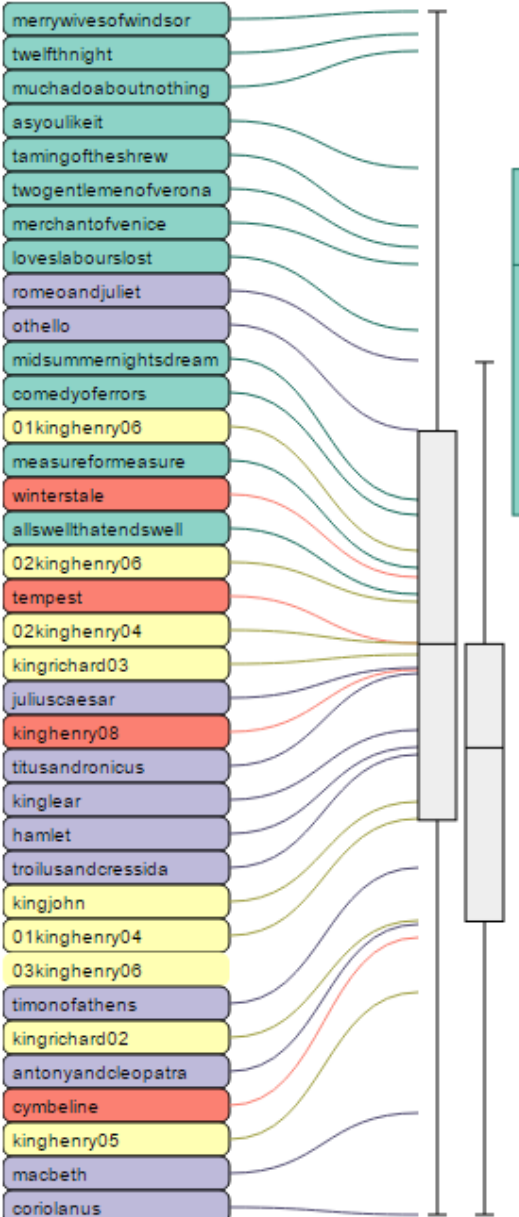
An SVM for Comedicness

Uses measurements of language

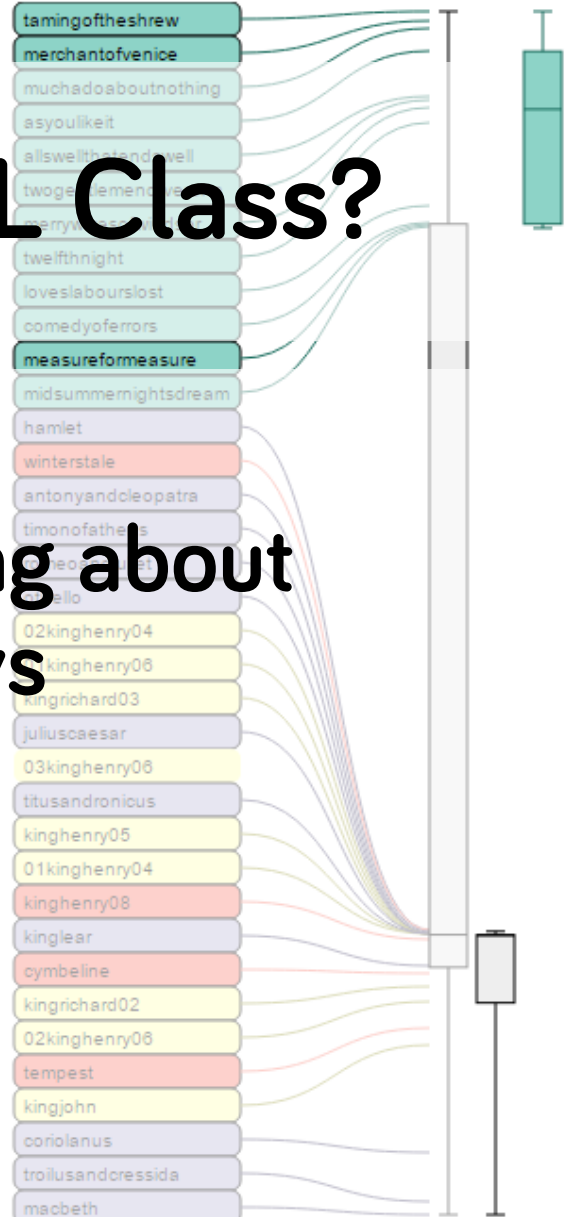


Did I fail my ML Class?

Or learn something about Shakespeare's plays



nth(0.96) mcc(0.76) margin(-0.19)
Docuscope.PredictedFuture:1
Docuscope.Inclusive:-1



nth(1) mcc(1) margin(0.58)
Docuscope.Precedent_Defending:-125.36
Docuscope.Apology:-45.61
Docuscope.Procedures:25.52
Docuscope.OpenQuery:10.66
and 19 others...

Simpler Functions



Easier to Understand



More likely
to lead to Theory



Less Expressive



Less Likely
to be Accurate
(but more likely to overfit)

Tradeoffs



Give the **user** control over the tradeoffs

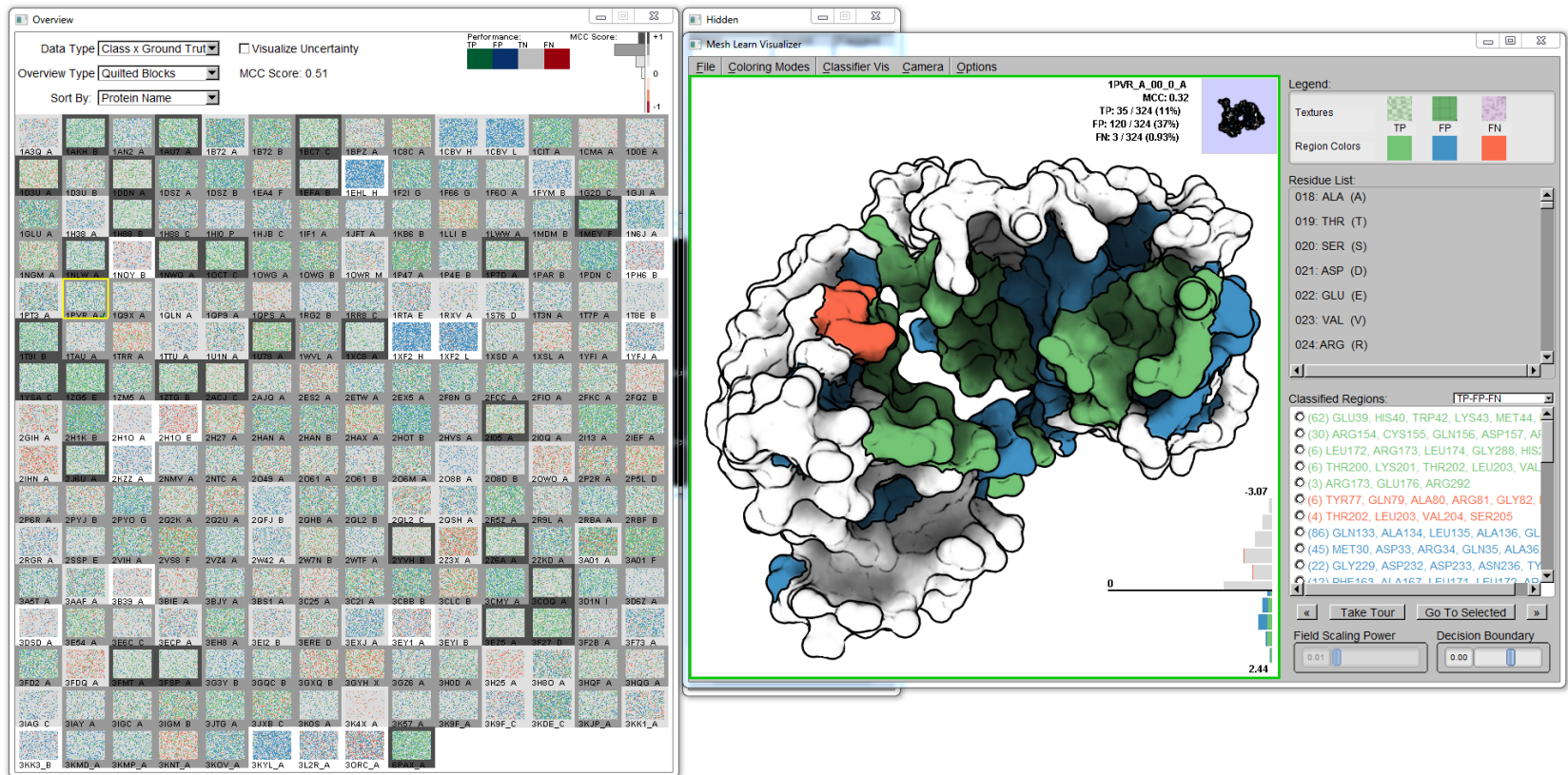
But how do we let them make informed choices?

Examine Model Outcomes

Look at validation experiments in detail
(not just as 1 number)

You **always** do good validation experiments, right?

Visual exploration of modeling validation experiments



Sarikaya, Albers, Mitchel and Gleicher. *Visualizing Validation of Protein Surface Classifiers*. Computer Graphics Forum. Proceedings EuroVis 2014

Who, Why, Where...

Who?

Stakeholders

Developers

Data Scientists

Domain Experts

Audience

Subjects

Why?

Reason for Wanting

Improve Performance

Build Theory

Extend/Characterize

Build Trust

Actionability

Convince

Where?

Phase of the **Process**

Identify Problem

Gather Data

Clean Data

Abstract

Design

Build

Evaluate

Interpret

Act

Disseminate

Who, Why, Where...

Who?

Stakeholders

Developers

Data Scientists

Domain Experts

Audience

Subjects

Why?

Reason for Wanting

Improve Performance

Build Theory

Extend/Characterize

Build Trust

Actionability

Convince

Where?

Phase of the **Process**

Identify Problem

Gather Data

Clean Data

Abstract

Design

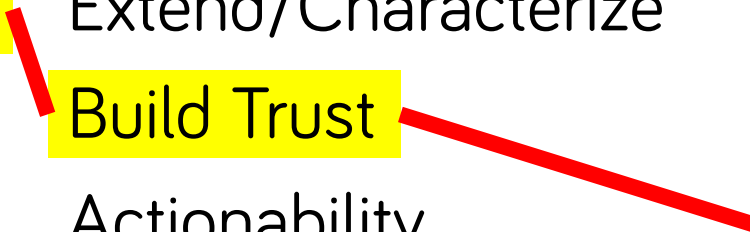
Build

Evaluate

Interpret

Act

Disseminate



Who, Why, Where...

Who?

Stakeholders

Developers

Data Scientists

Domain Experts

Audience

Subjects

Why?

Reason for Wanting

Improve Performance

Build Theory

Extend/Characterize

Build Trust

Actionability

Convince

Where?

Phase of the **Process**

Identify Problem

Gather Data

Clean Data

Abstract

Design

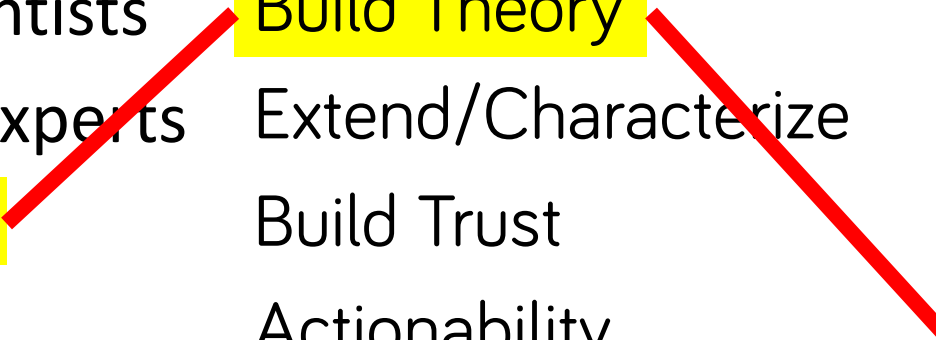
Build

Evaluate

Interpret

Act

Disseminate



More?

(probably not – this is already too long)

Help Data Tasks

Data cleanup

Model Building

End-user decision making

(and more)

Approaches

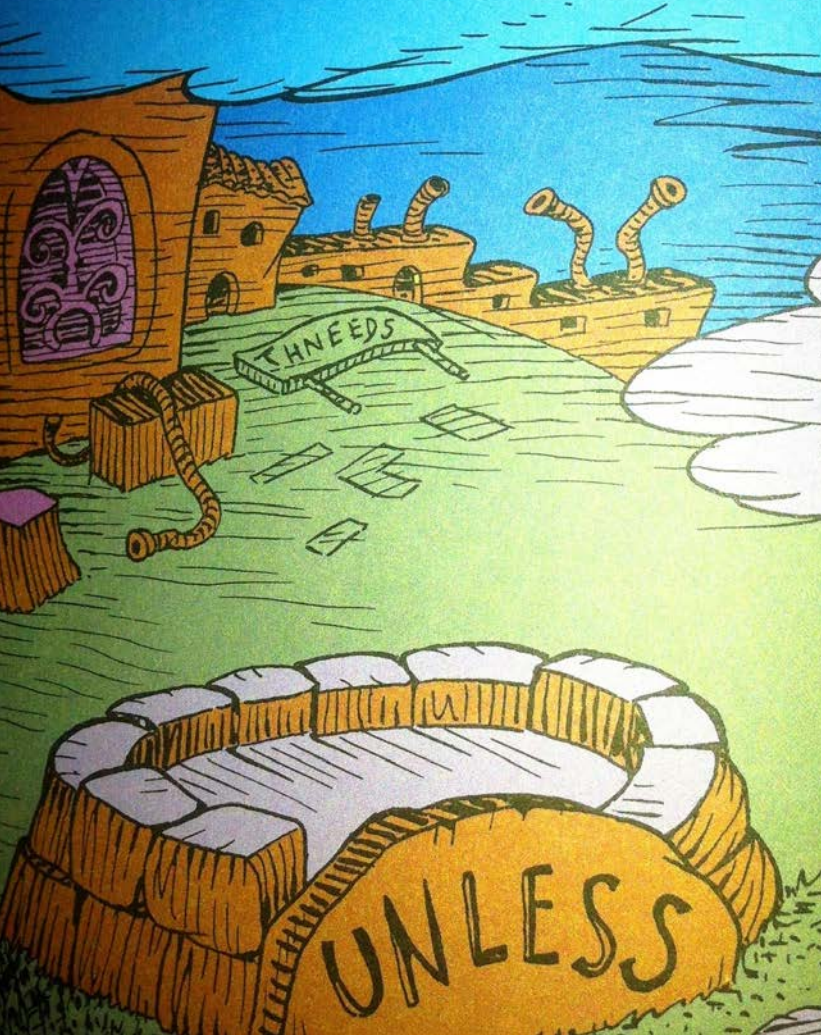
New model types

Interactive modeling tools

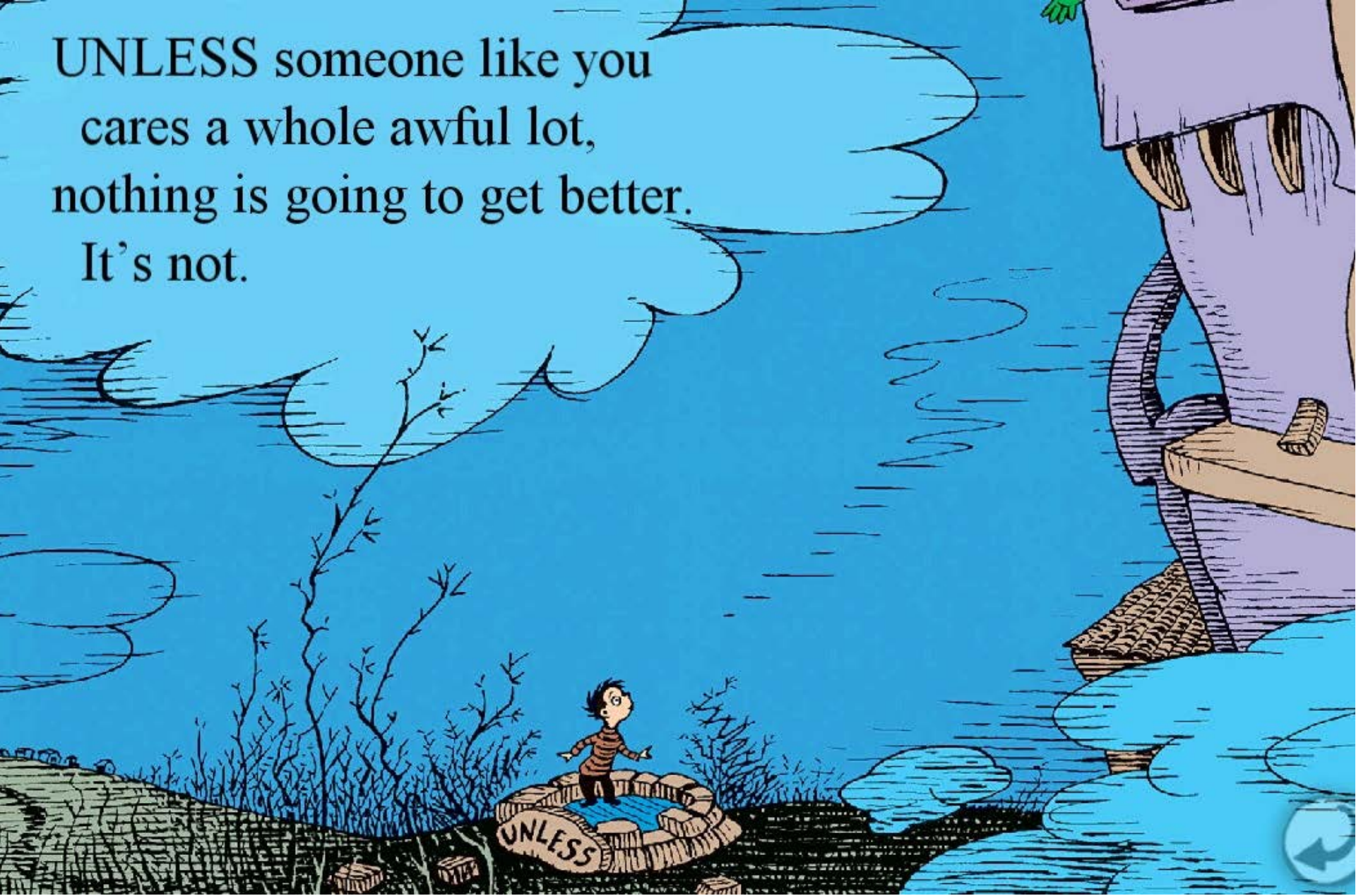
New outcome explorations

(and more)

And all that the Lorax left here in this mess
was a small pile of rocks, with the one word...
"UNLESS."
Whatever *that* meant, well, I just couldn't guess.



UNLESS someone like you
cares a whole awful lot,
nothing is going to get better.
It's not.



Thanks!

To you for listening.

To the organizers for inviting me

To my students and collaborators.

To the NSF, NIH and Mellon Foundation for funding.

Towards Comprehensible Modeling

Michael Gleicher

University of Wisconsin Madison

gleicher@cs.wisc.edu

