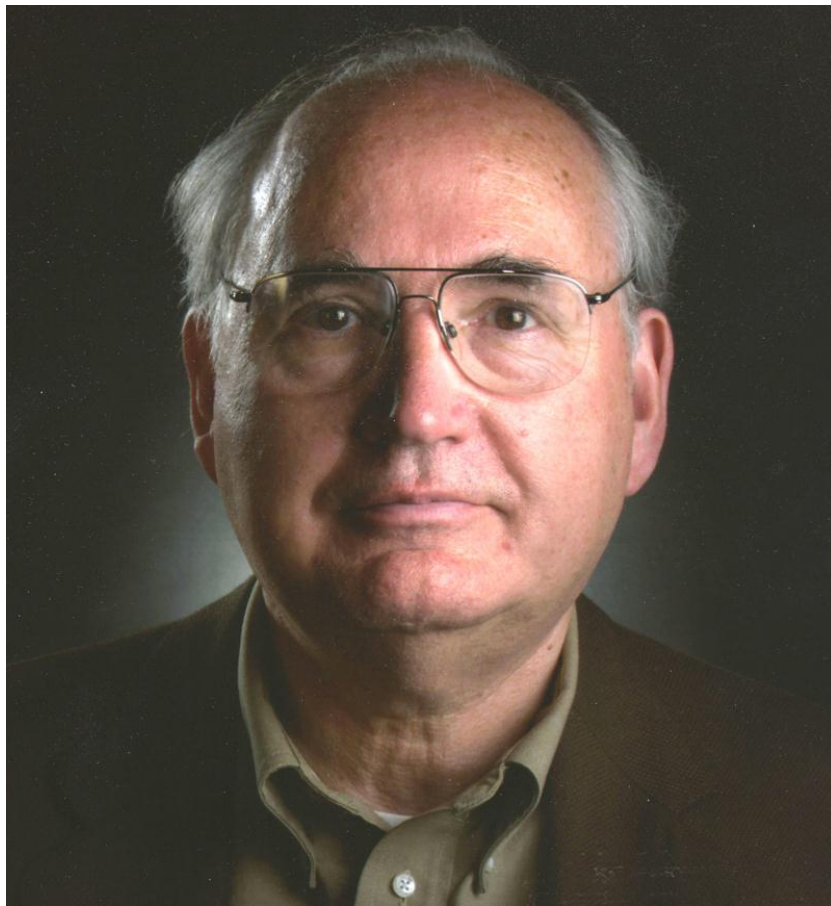


Neither Basic Nor Applied

Lessons From a Computer Research Journey

Peter Lee

Microsoft Research Redmond



Dana Scott

$F(n) = \text{if } n == 0 \text{ then return } 1 \text{ else return } n * F(n-1)$

F = "the factorial function"
is a solution...

...but it is also a solution for

$F(n) = \text{if } n == 0 \text{ then return } 1 \text{ else return } F(n+1) / (n+1)$

“If you think that beauty is so important, maybe you should be in the School of Art instead of the Computer Science Department.”

*— a former CMU CS Department Head,
speaking to Peter Lee, Assistant Professor*

“Things like even software verification, this has been the Holy Grail of computer science for decades but now in some very key areas, for example, driver verification we’re building ***tools that can do actual proof about the software*** and how it works in order to guarantee the reliability.”

– *Bill Gates, April 18, 2002*

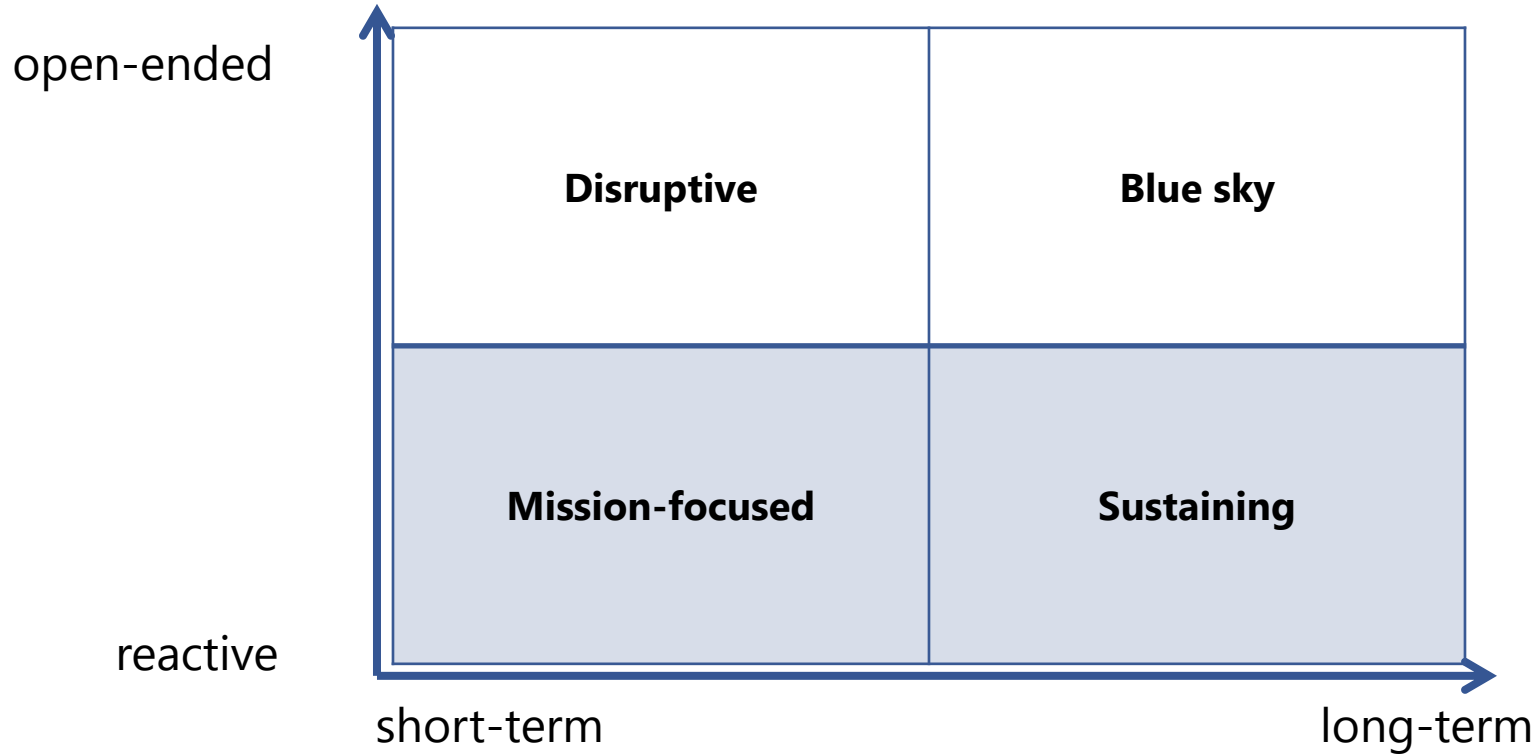




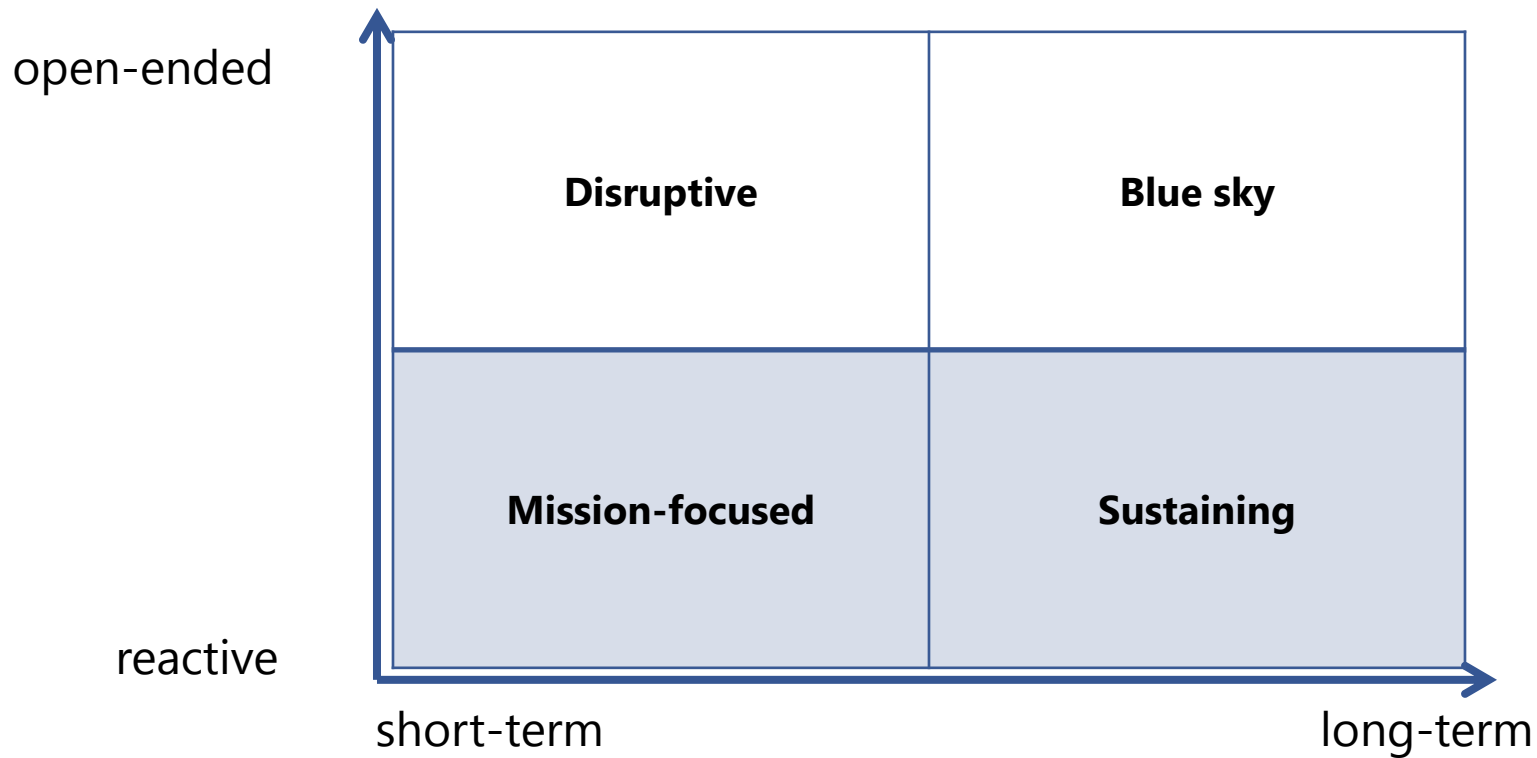
$F(n) =$
 if n : **blue-sky** return 1
 else return $n * F(n-1)$

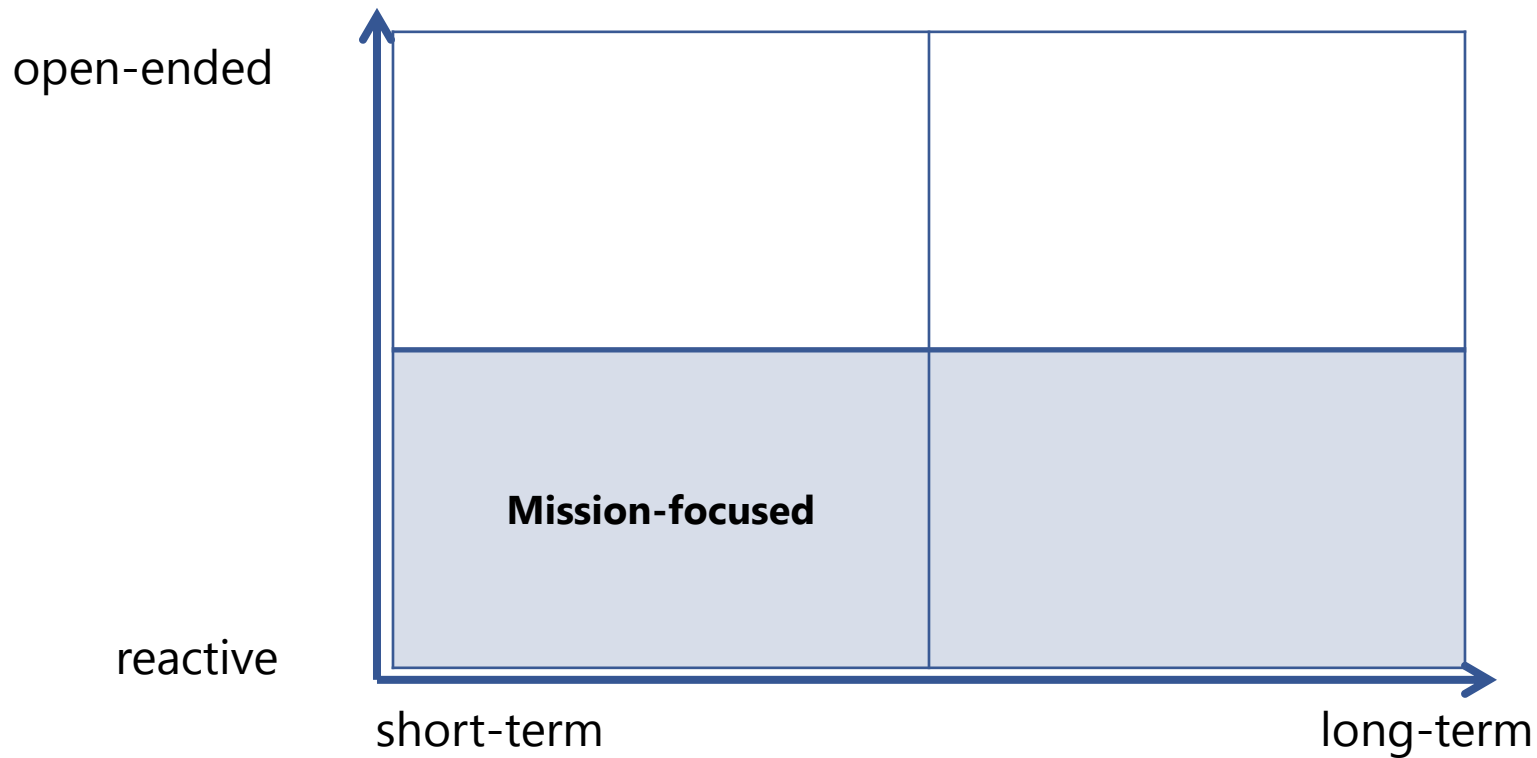


“Mudge’s quadrants”



Microsoft®
Research







SPECIAL ARTICLE

[◀ Previous](#)

Volume 360:1418-1428

[April 2, 2009](#)

Number 14

[Next ▶](#)

Rehospitalizations among Patients in the Medicare Fee-for-Service Program

Stephen F. Jencks, M.D., M.P.H., Mark V. Williams, M.D., and Eric A. Coleman, M.D., M.P.H.

ABSTRACT

Background Reducing rehospitalizations is a way to improve quality of care and reduce the frequency and potential costs of unnecessary changes in care.

Methods We analyzed rehospitalization among Medicare patients and to characterize

Results Almost one fifth

discharged from a hospital were rehospitalized within 30 days, and 34.0% were rehospitalized within 90 days; 67.1% of patients who had been discharged with medical conditions and

- ~20% within 30 days

- ~35% in 90 days

- **Estimated cost to Medicare in 2004: \$17.4 billion**

Predictive analytics via Machine Learning

Weight	Feature description	Frequency
0.68398	Dx0->2 = Excessive vomiting in pregnancy	0.31%
0.61306	Dx3->2 = Personal history of malignant neoplasm	0.28%
0.58281	Dx0->2 = Heart failure	0.30%
0.56708	Dx0->1 = Nephritis, nephrotic syndrome, and nephrosis	0.09%
0.56649	Dx3->2 = Heart failure	0.28%
0.54663	Complaint sentence contains "suicidal"	0.17%
0.48415	Dx1->2 = Disorders of function of stomach	0.07%
0.47257	Dx5->0 = Diseases Of The Genitourinary System	0.15%
0.46136	Dx0->2 = Chronic airway obstruction, not elsewhere classified	0.10%
0.44555	Dx4->2 = Depressive disorder, not elsewhere classified	0.10%
0.44257	Stayed 14 hours in the ER	0.10%
0.43890	Dx0->1 = Other psychoses	0.32%
0.43513	Dx0->0 = Diseases Of The Blood And Blood-Forming Organs	0.46%
0.42582	Complaint sentence contains "dialysis"	0.19%
0.41888	Dx0->2 = Depressive disorder, not elsewhere classified	0.27%
0.41302	Dx1->1 = Nephritis, nephrotic syndrome, and nephrosis	0.29%
0.38506	Complaint sentence contains "fluid"	0.10%
0.37474	69 < Age	9.22%

Learning to Rank

Chris J. C. Burges
Microsoft Research



learning to rank



Web

Web

Videos

More ▾

RELATED SEARCHES

- Learning to Rank Challenge
- Yahoo! Learning to Rank Challenge
- Yahoo! Learning to Rank
- 5Vbm25
- RankNet
- Tie Yan Liu
- Ndcg Search
- Lambda Rank

SEARCH HISTORY

Search more to see your history

ALL RESULTS

1-10 of 106,000,000 results · [Advanced](#)

[Learning to rank - Wikipedia, the free encyclopedia](#)

[Applications](#) · [Feature vectors](#) · [Evaluation measures](#) · [Approaches](#)

Learning to rank or **machine-learned ranking** (MLR) is a type of supervised or semi-supervised machine **learning** problem in which the goal is to automatically construct a **ranking** ...

en.wikipedia.org/wiki/Learning_to_rank

[Yahoo! Learning to Rank Challenge](#)

Learning to Rank Challenge is closed! Close competition, innovative ideas, and fierce determination were some of the highlights of the first ever Yahoo!

learningtorankchallenge.yahoo.com

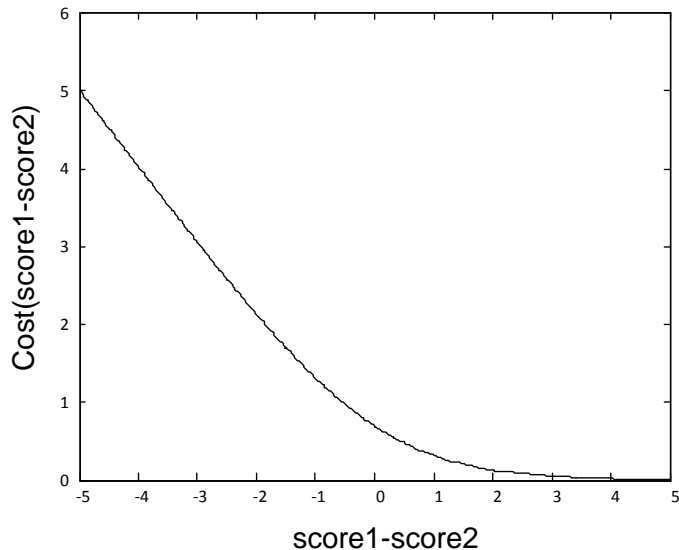
[Learning to Rank using Gradient Descent](#)

Learning to Rank using Gradient Descent Keywords: **ranking**, gradient descent, neural networks, probabilistic cost functions, internetsearch Chris Burges cburges@microsoft.com ...

research.microsoft.com/en-us/um/people/cburges/papers/ICML_ranking.pdf · PDF file

How LambdaRank is Trained

● Query: "107.7 the end"



	Model Score
	0.23
	0.21
	0.19
	0.13
	0.01

$$y_1 = -0.525$$

$$y_2 = -0.040$$

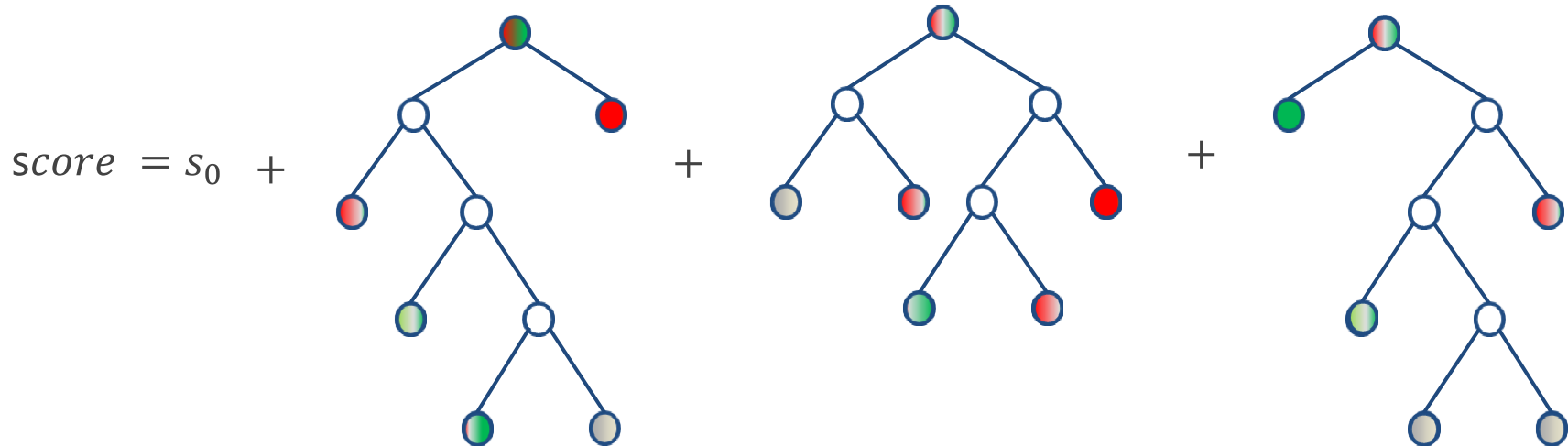
$$y_3 = 0.038$$

$$y_4 = 0.551$$

$$y_5 = -0.023$$

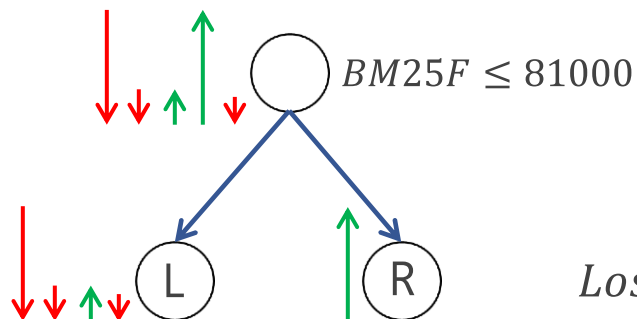
Building an ensemble of regression trees

Rating	Url	Score	Force
Good	http://en.wikipedia.org/wiki/107.7_The_End	0.20	-0.09
Perfect	http://www.1077theend.com	0.19	0.37
Good	http://www.myspace.com/1077theend	0.19	0.04
Bad	http://www.thewolf.co.uk/	0.17	-0.35
Fair	http://en.wikipedia.org/wiki/The_End	0.01	0.10



How is a split computed?

Rating	Url	BM25F	Count of query in body	Anchor text matches query	Function(click frequency)
Bad	http://www.thewolf.co.uk/	55370	13	0	0
Good	http://en.wikipedia.org/wiki/107.7_The_End	81000	136	2	0
Good	http://www.myspace.com/1077theend	80981	0	0	0
Perfect	http://www.1077theend.com	81023	60	25286	251
Fair	http://en.wikipedia.org/wiki/The_End	80984	156	2049	0



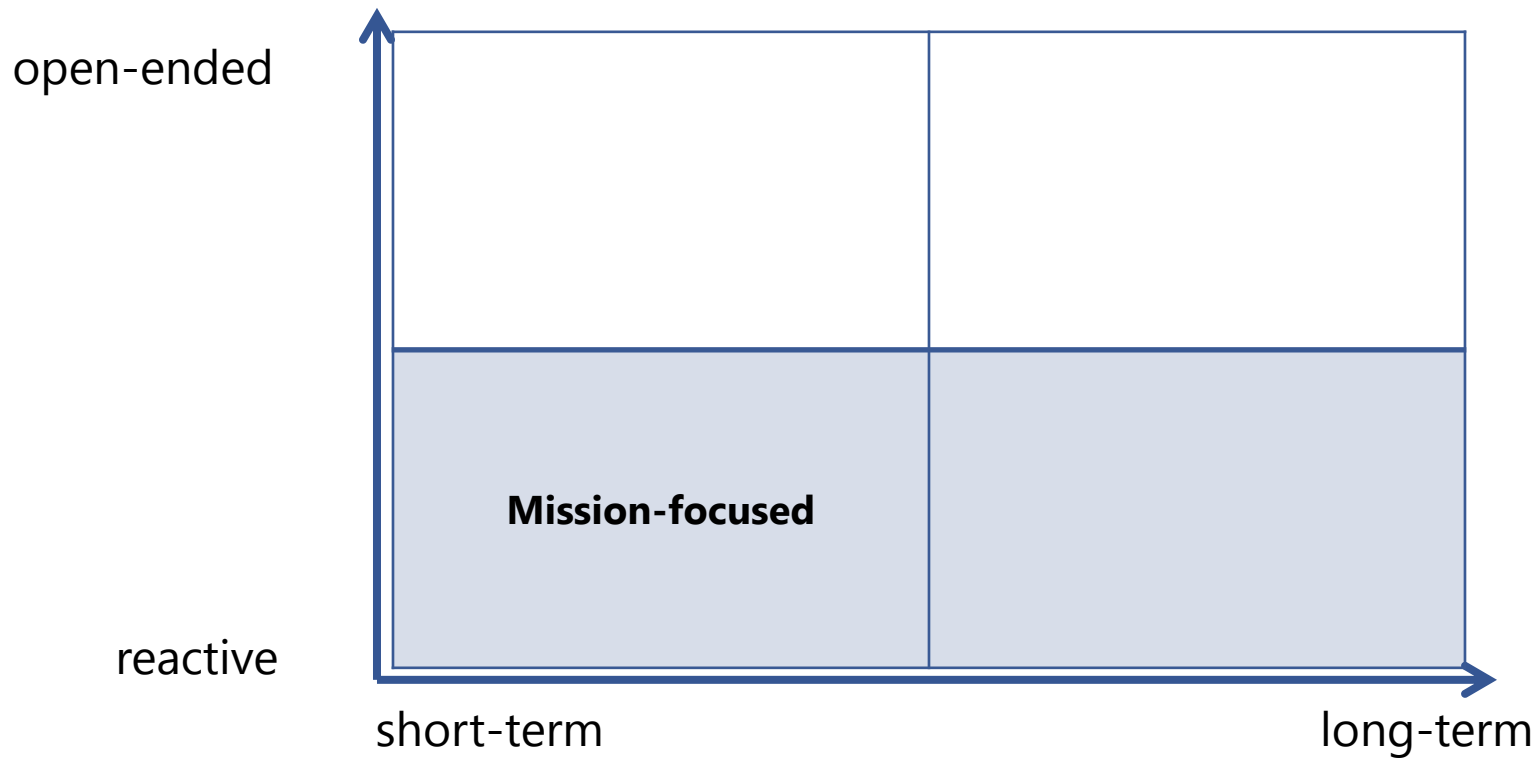
$$Loss = \sum_i (y_i - \bar{y})^2$$

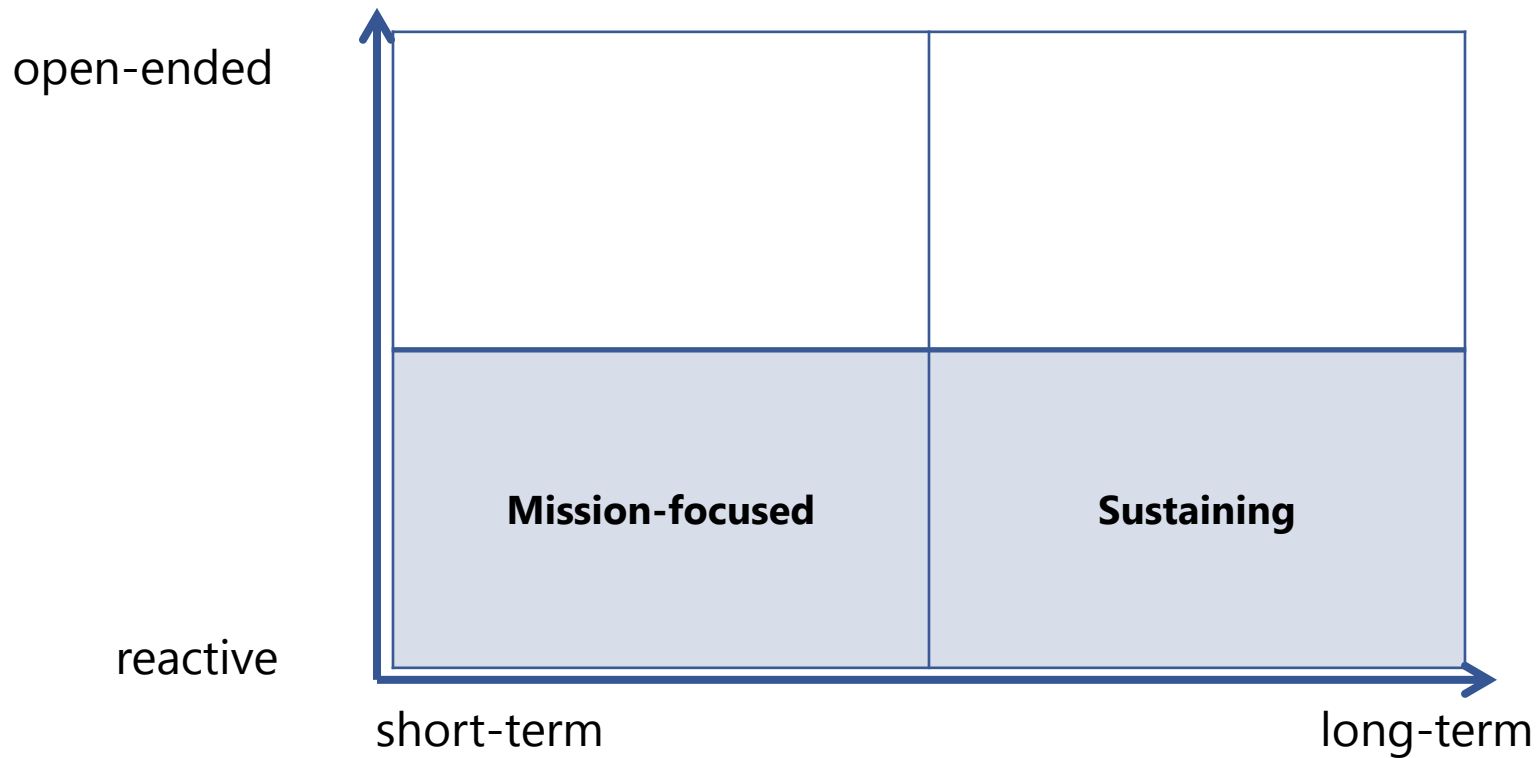
$$Loss = \sum_{i \in L} (y_i - \bar{y}_L)^2 + \sum_{i \in R} (y_i - \bar{y}_R)^2$$

FastRank: Algorithms \leftrightarrow Engineering

Minimizing the loss is equivalent to maximizing $\frac{1}{|L|} (\sum_{i \in L} y_i)^2 + \frac{1}{|R|} (\sum_{i \in R} y_i)^2$

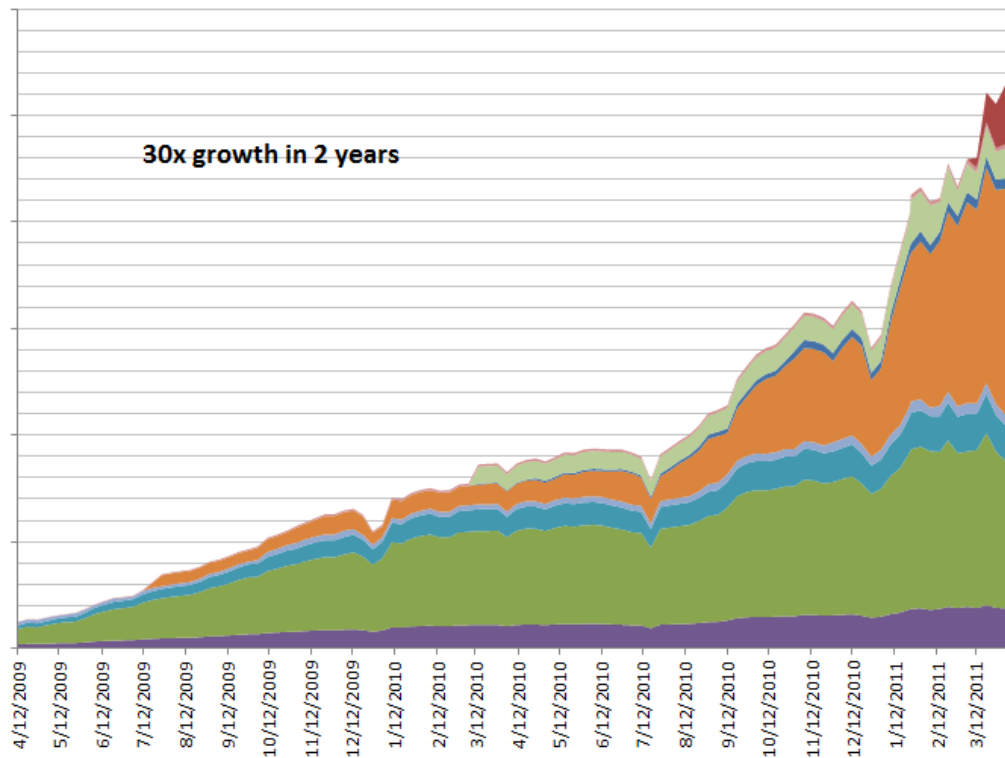
We only need *one linear scan* of the data (in a given node) to compute this.





Microsoft Translator

Daily Translation Requests



Whitebox Fuzzing

David Molnar
Microsoft Research

Problem: Security Bugs in File Parsers

Ongoing challenge for Microsoft ecosystem



Hundreds of file formats
supported in Windows, Office, etc.

Microsoft Security Bulletin Summary for July 2011

Every security patch costs Microsoft
alone **one million dollars.**



All groups do “fuzz testing,”
mandated by the Security
Development Lifecycle.

Traditional random fuzz testing
can't catch this bug:

```
int obscure(int x, int y) {  
    if (x==hash(y)) error();  
    return 0;  
}
```


Approach: "Whitebox" Fuzz Testing

State of the art solving + industrial strength binary analysis

```
void top(char input[4])
```

```
{
```

```
int cnt = 0;
```

```
if (input[0] == 'b') cnt++;
```

```
if (input[1] == 'a') cnt++;
```

```
if (input[2] == 'd') cnt++;
```

```
if (input[3] == '!') cnt++;
```

```
if (cnt >= 3) crash();
```

```
}
```

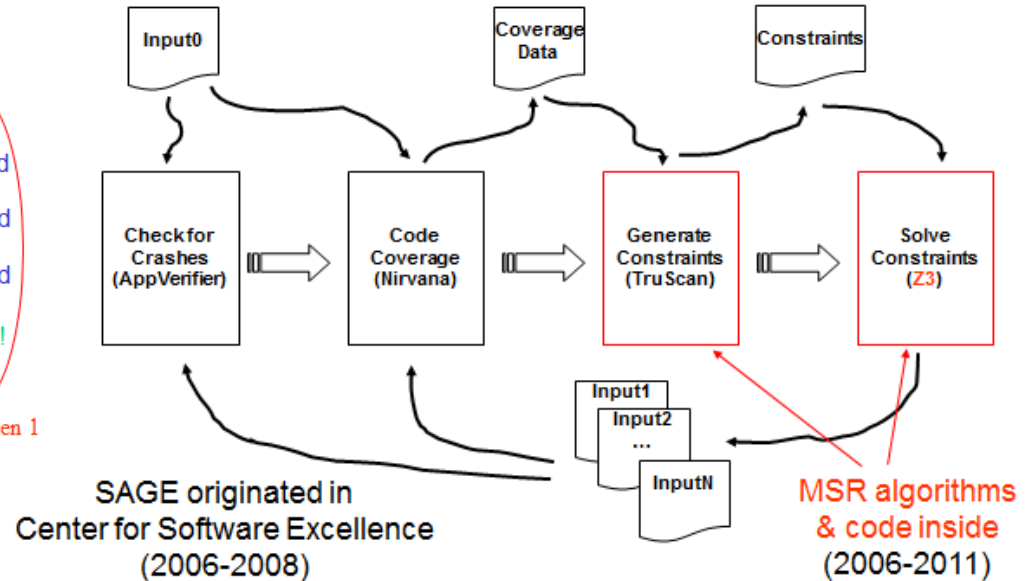
input = "good"

Path constraint:



Negate each constraint in path constraint

Solve new constraint → new input



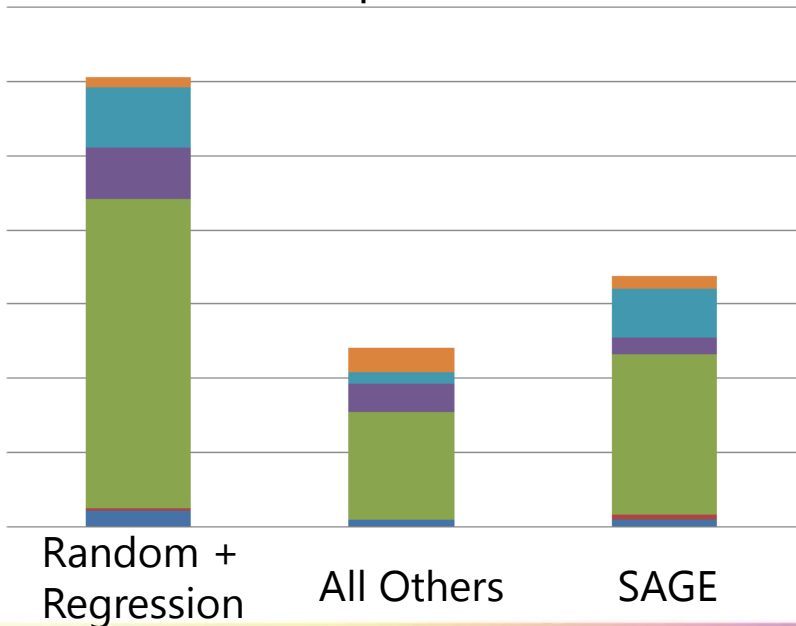
DEMO

SAGE and SAGAN

Impact: Change Security Testing

Largest deployment of solvers in the world

Fuzzing bugs found in Win7 "WEX"
over 100s of file parsers:

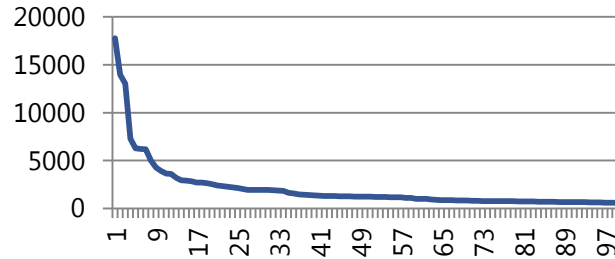


- Run in every Win8 milestone
- Ongoing engagement with Office and Microsoft Security Engineering Center
- **200+** machine years
- **One billion+** constraints
- Big **thank-you** to our partners across Microsoft!

Research: Feedback From Scale

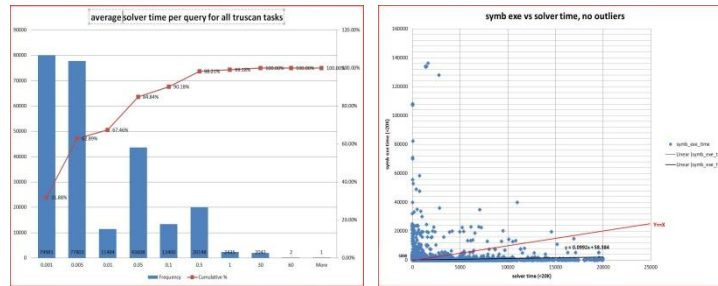
Any test anywhere helps all future tests everywhere!

How much **sharing** between symbolic execution of **different** programs run on Windows?



Most common branch appears **17761** times out of **290430** symbolic executions. Motivates **symbolic summaries built up over time**.

How does the Z3 solver perform on **constraints** arising from **real code**?



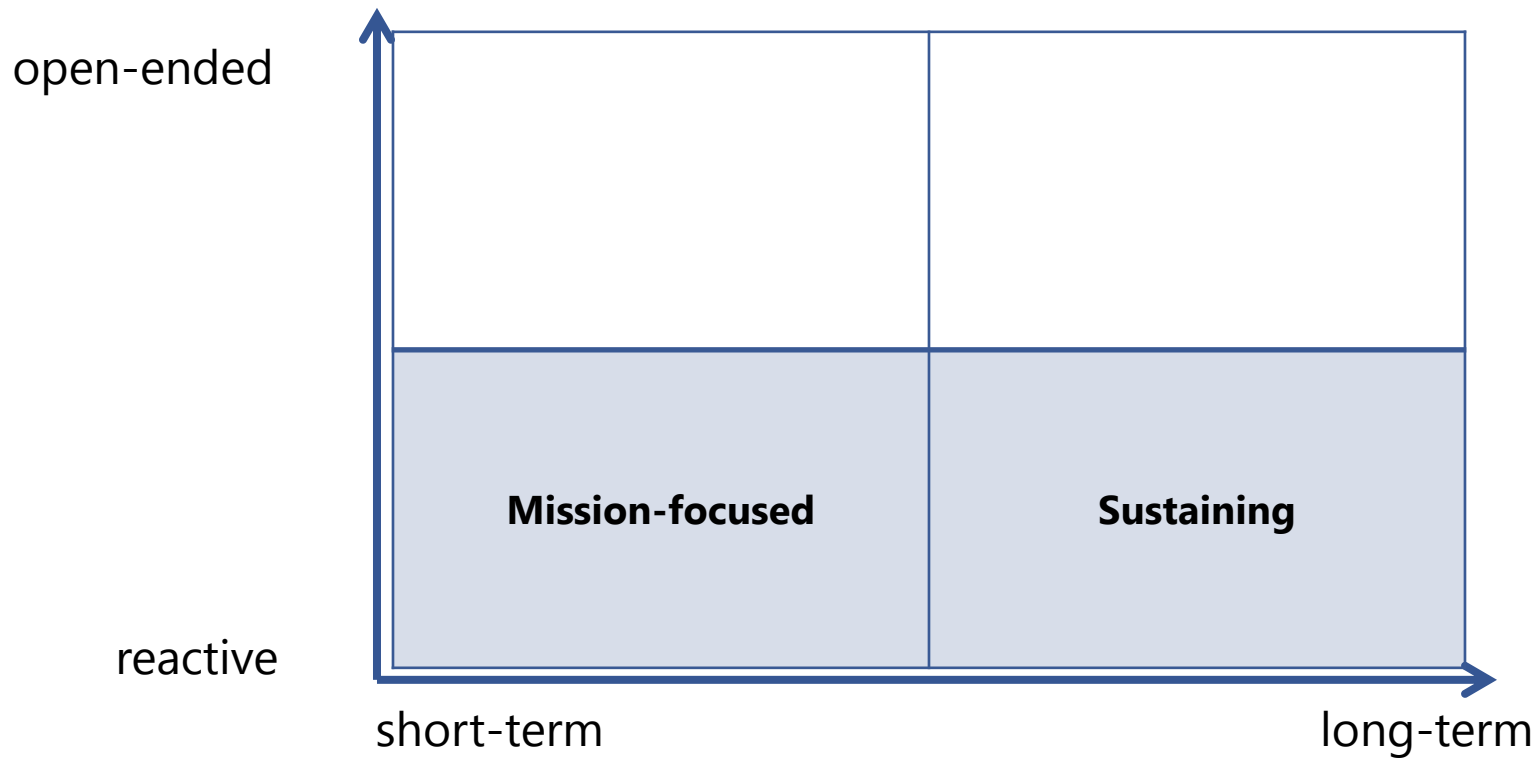
90.18% of Z3 queries solved in **0.1 seconds or less**. Solving time still dominates! Tells us where to focus Z3.

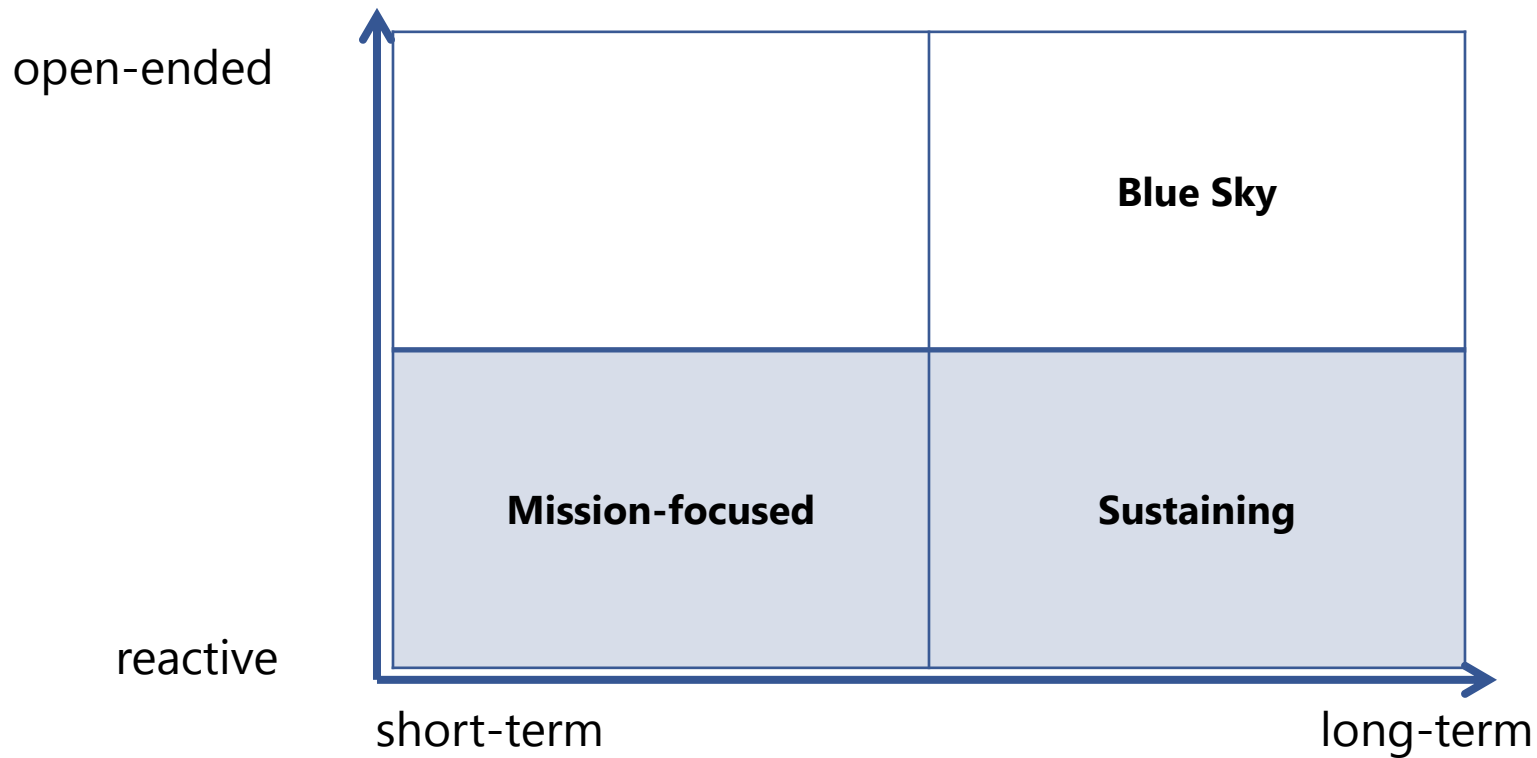
Leverage data collection to create **virtuous cycle** that keeps us ahead of the competition!

Answer questions and pursue directions **impractical without scale**.

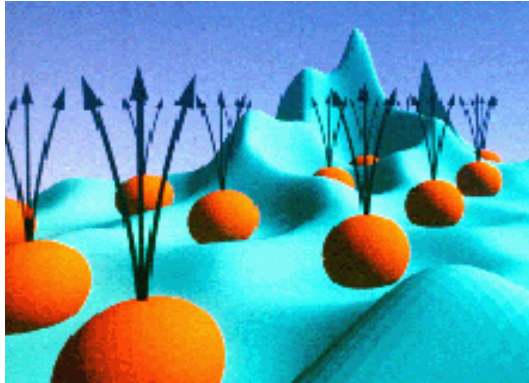
Thanks to all SAGE contributors!

- **MSR**: Ella Bounimova, Patrice Godefroid, David Molnar
(+ our managers for their support! 😊)
- **CSE**: Michael Levin, Chris Marsh, Lei Fang, Stuart de Jong,...
- **Interns** : Dennis Jeffries (06), David Molnar (07), Adam Kiezun (07), Bassem Elkarablieh (08), Marius Nita (08) , Cindy Rubio-Gonzalez (08,09), Johannes Kinder (09), Daniel Luchaup (10), Mehdi Bouaziz (11),...
- **Z3 (MSR)**: Nikolaj Bjorner, Leonardo de Moura,...
- **Windows**: Nick Bartmon, Eric Douglas, Dustin Duran, Elmar Langholz , Isaac Sheldon, Dave Weston,...
 - Win8 TruScan support: Evan Tice, David Grant,...
- **Office**: Tom Gallagher, Eric Jarvi, Octavian Timofte,...
- **MSEC**: Dan Margolis, Matt Miller, Lars Opstad, Jason Shirk,...
- **SAGE users all across Microsoft!**
- Download SAGE (Microsoft only): <http://sharepoint/sites/SAGE>

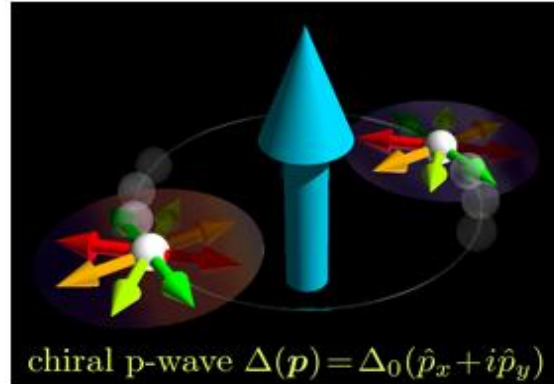




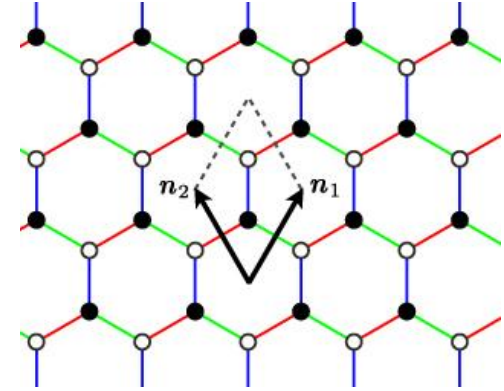
Topological Quantum Systems



Fractional Quantum Hall States



Topological Superconductors



Spin Systems

Topological devices are immune to local errors,
Thus relieving much of the burden of error correction.

Programming on the Phone

Nikolai Tillmann
Microsoft Corporation

Program smartphones directly!



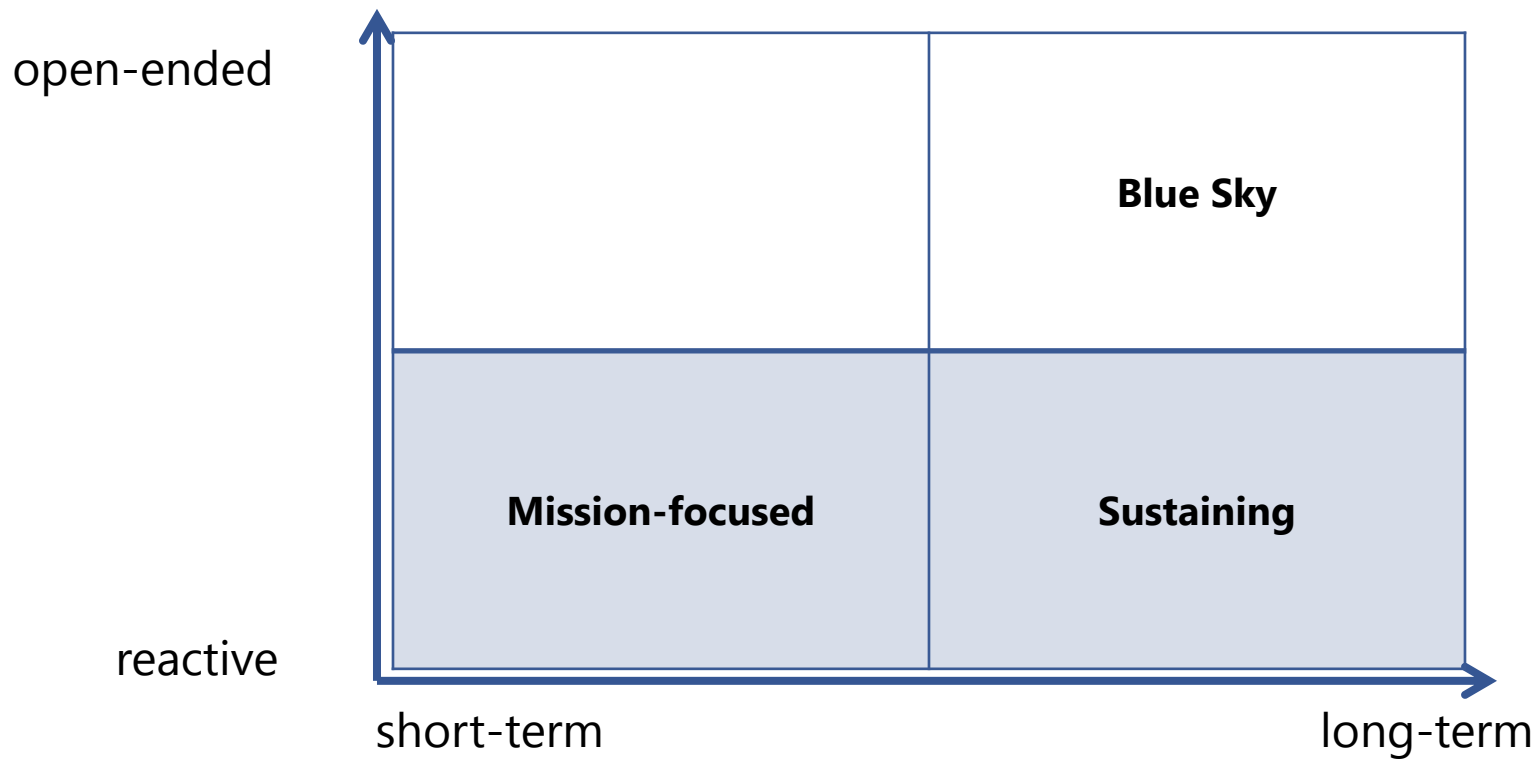
- In 2010, smartphones outsold PCs
- Today's smartphones are more powerful than PCs from 2000
- How do we use them?

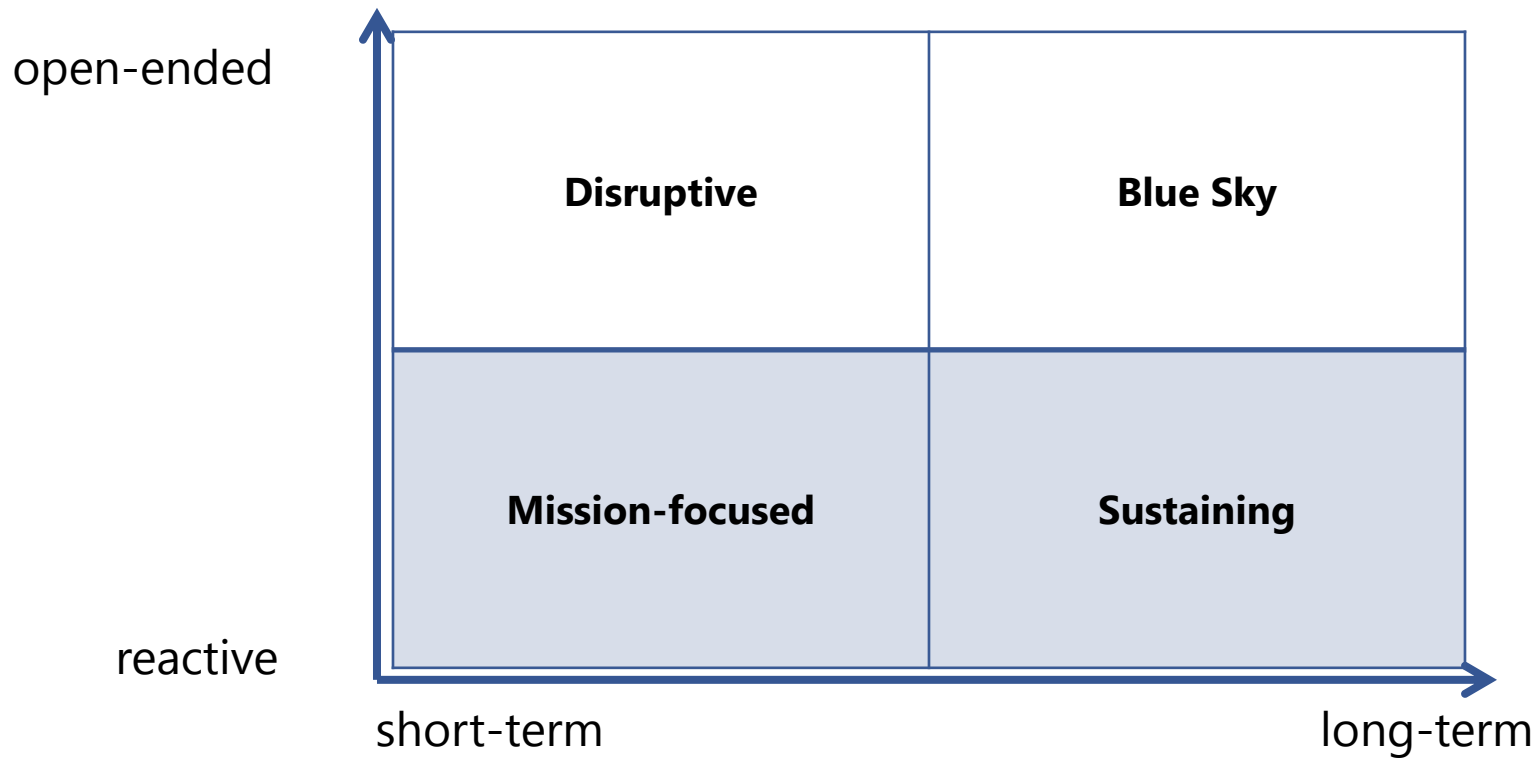
DEMO

Programming on the Phone

Programming directly on the phone...

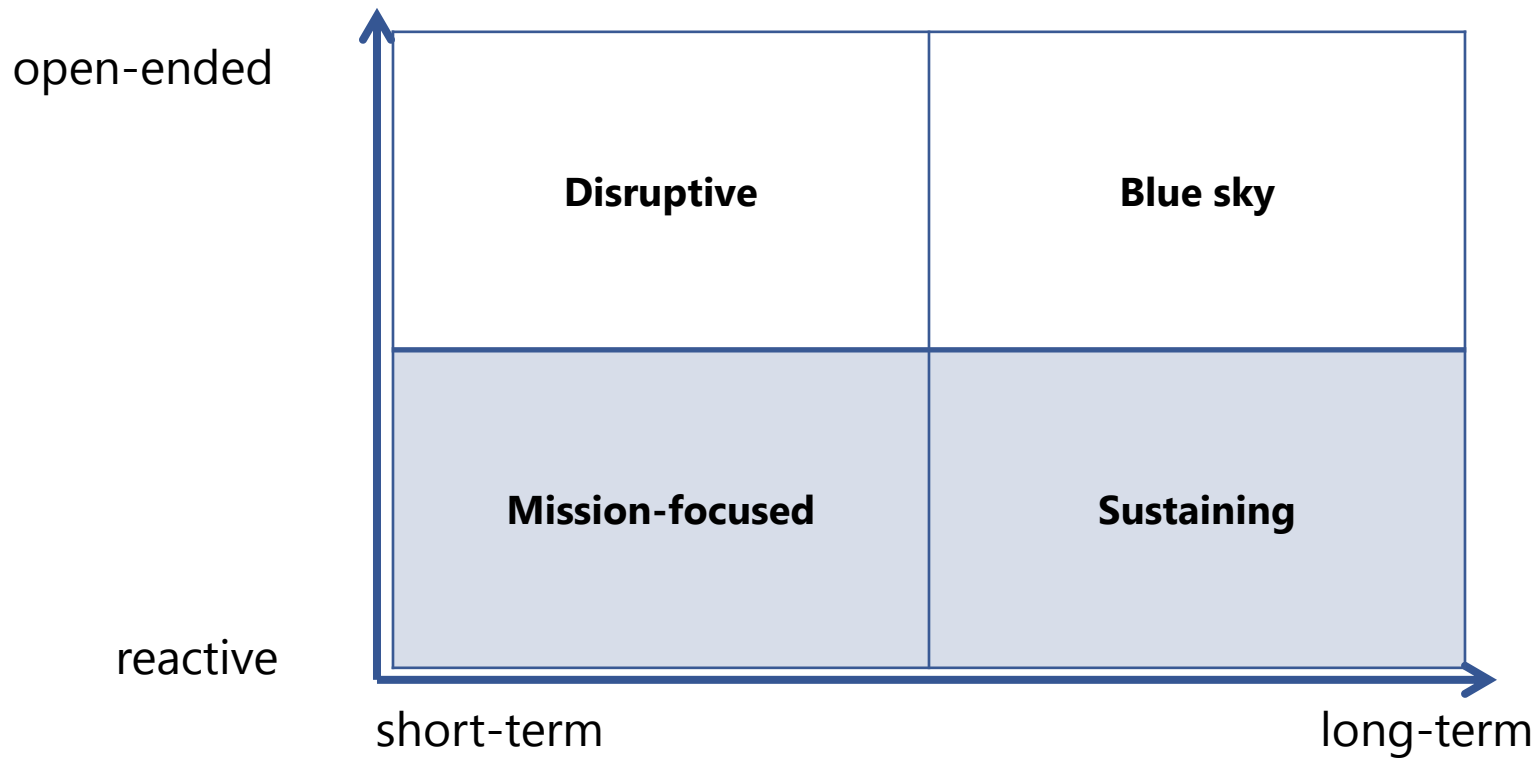
- Is popular
 - Top 200 of Windows Phone Marketplace; top 10 of Productivity apps
 - Rated 4.7/5 by users
- Is great for teaching
 - Engaging programming experience
 - Using personal device and data (songs, pictures, etc.)
- Embraces new reality
 - Language and editor optimized for touchscreen
 - First-class access to sensors and services
 - Sharing and learning with friends in the cloud





Audio for Kinect

Ivan Tashev
Microsoft Research



Computing research...

...has incredible diversity,

and rarely is exclusively “basic” or “applied”.

Thank you!

petelee@microsoft.com



Microsoft Research
Faculty Summit

© 2011 Microsoft Corporation. All rights reserved. Microsoft, Windows, Windows Vista and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries.

The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation.

MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.

Microsoft Research

FacultySummit



FUTURE WORLD

2011 ← → 2031