



LADIES AND GENTLEMEN, START YOUR ENGINE!

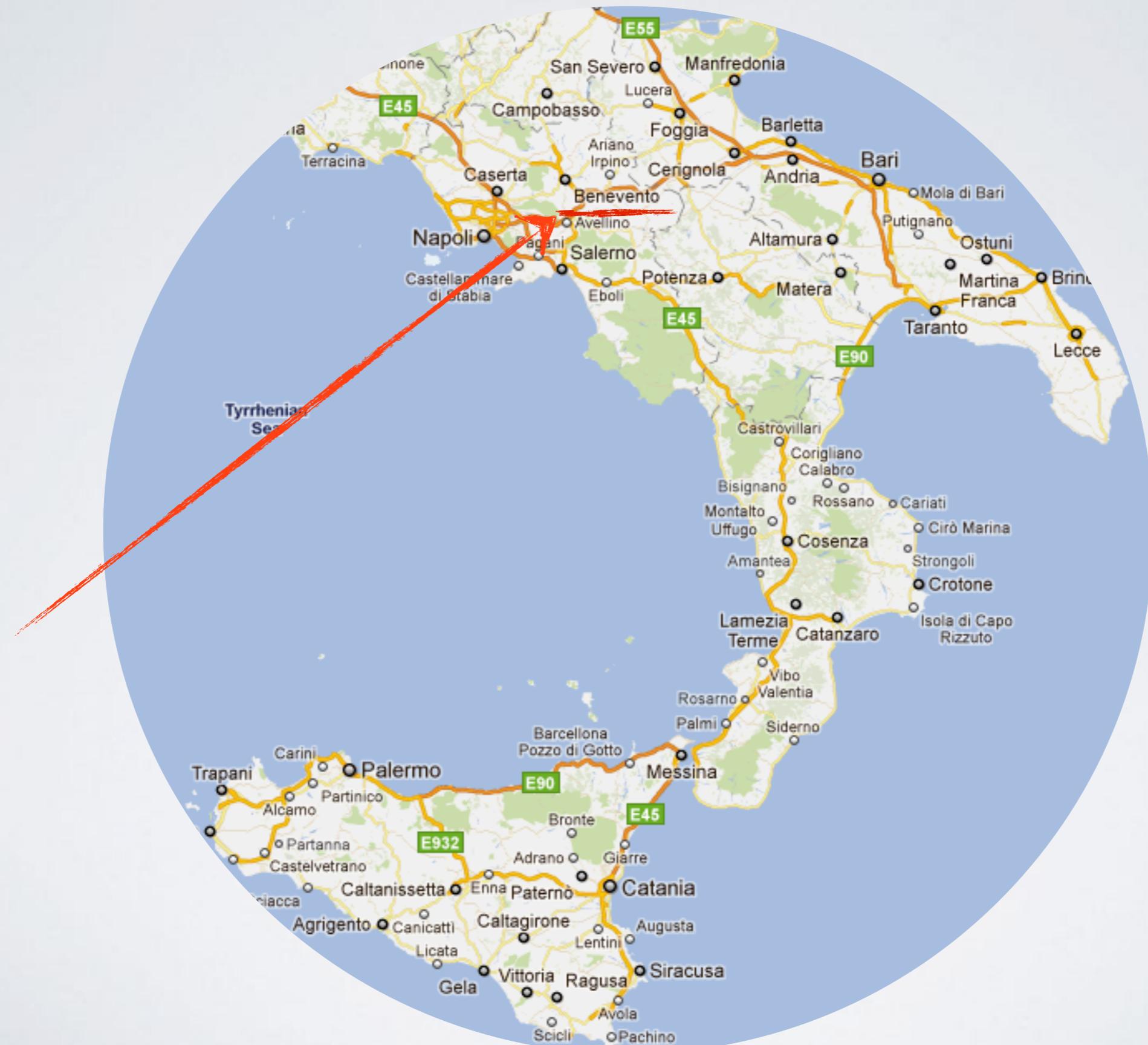
Being Successful with your PhD
Without Crashing at the First Corner

Massimiliano Di Penta

University of Sannio, Italy

dipenta@unisannio.it, [@maxdipenta](https://twitter.com/maxdipenta)

UNIVERSITY OF...WHAT?



ABOUT ME

Teaching Advanced Software Engineering (Master)
+ an introductory course (Bachelor)
+ empirical software engineering (PhD)

Chair of the computer science engineering
Bachelor and Master program

I review A LOT

MAIN RESEARCH INTERESTS

Software evolution

Software analytics

Empirical software engineering

Software testing

OTHER INTERESTS



Monza, Sep 2015
(after ESEC-FSE'15)

OUTLINE



Conducting
your Research

Contingency
Plans

Managing
your PhD

Next Steps

LET'S START!



CONDUCTING YOUR RESEARCH



CHOOSING YOUR PHD TOPIC

Often you spend the first months just doing this...

THE TOPIC

- (Still) relevant in research and practice
- You feel there is room to contribute
- Your research group has competences
- You like it

GOOD RESEARCH

Solves a real problem practitioners have

Builds something useful for future researchers

Makes observations from which others can learn

PITFALLS (IMHO)



WE HAVE A SOLUTION, LET'S
FIND A PROBLEM FOR IT



SALAMI PAPERS



RUSHING TO THE NEXT DEADLINE



HOWEVER....

RACE STARTS AT 2PM



Conference submissions close at Midnight AOE
(2pm CET)

PLANNING IS ESSENTIAL

SHANGHAI - PIT STOP SUMMARY

AVAILABLE
TYRES



WHITE MEDIUM



YELLOW SOFT



RED SUPERSOFT



GREEN INTERMEDIATE



BLUE WET



PARK FERME' (CLOSED PARK)



Refrain to perform massive changes

- New computations
- Major manuscript reorganization

BE PICKY

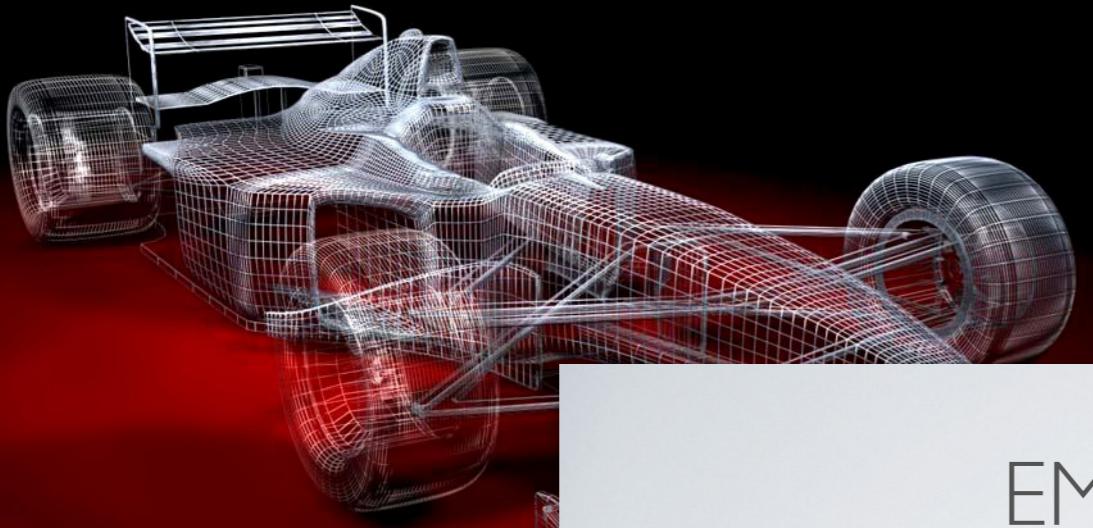
Proof read, proof read, proof read

Avoid typoz

Polish any aspect of the paper (bib, figures, tables)

DIFFERENT CONTRIBUTIONS

THEORETICAL



PRACTICAL

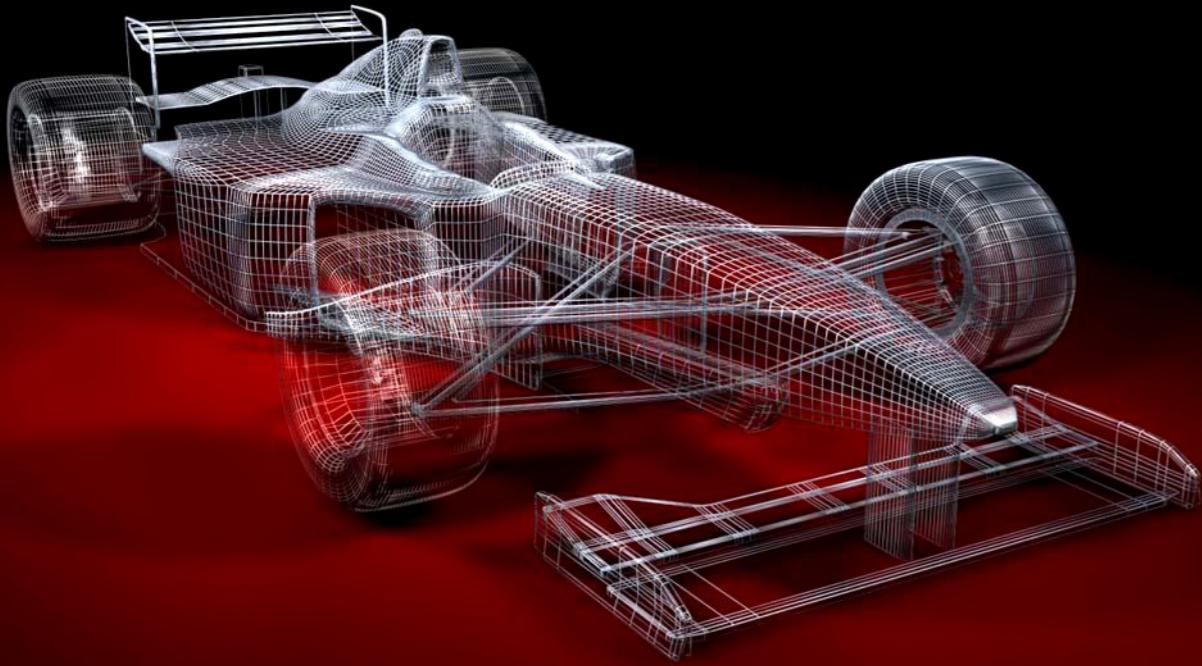


EMPIRICAL



DIFFERENT CONTRIBUTIONS

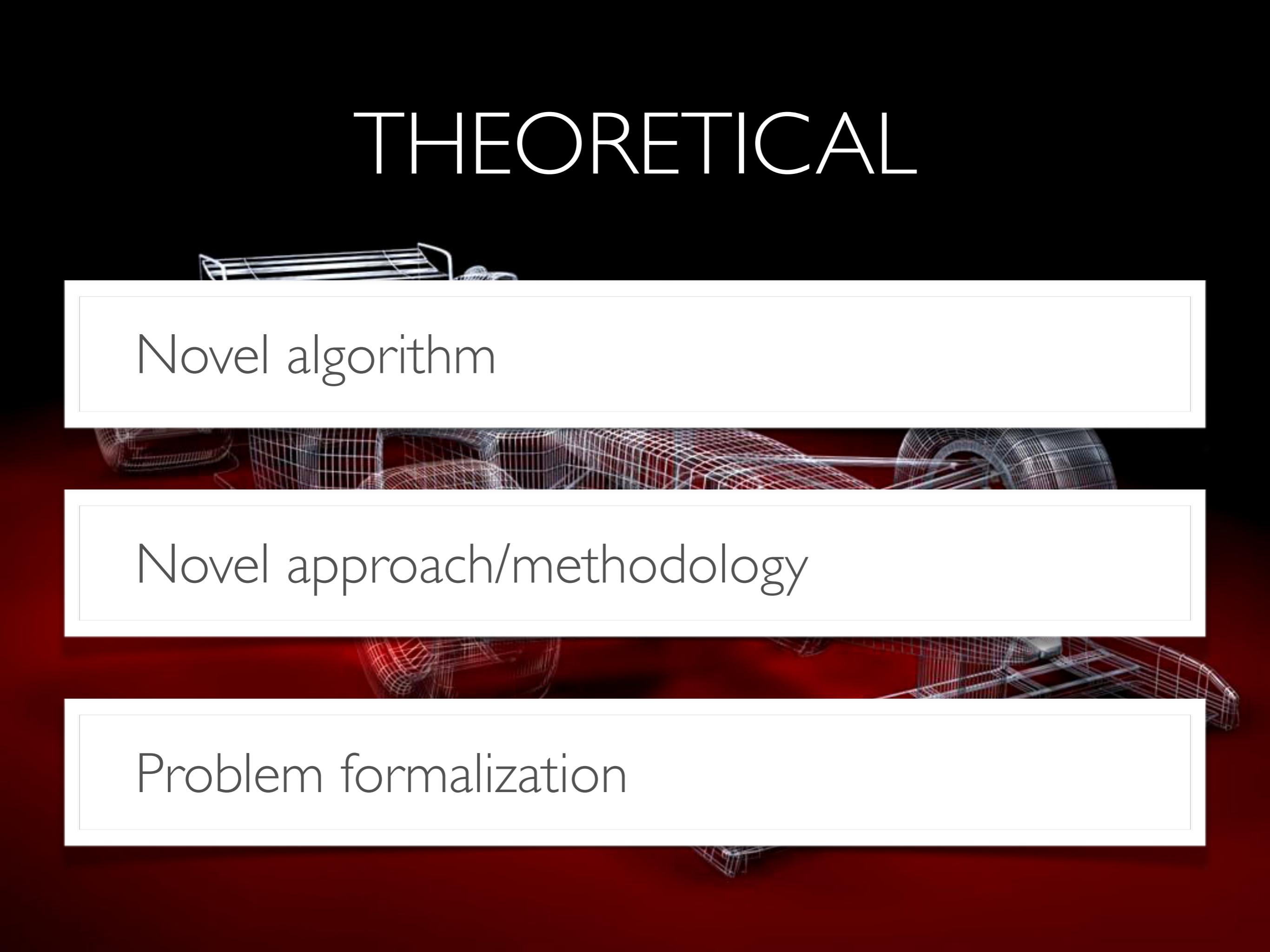
THEORETICAL



PRACTICAL



THEORETICAL



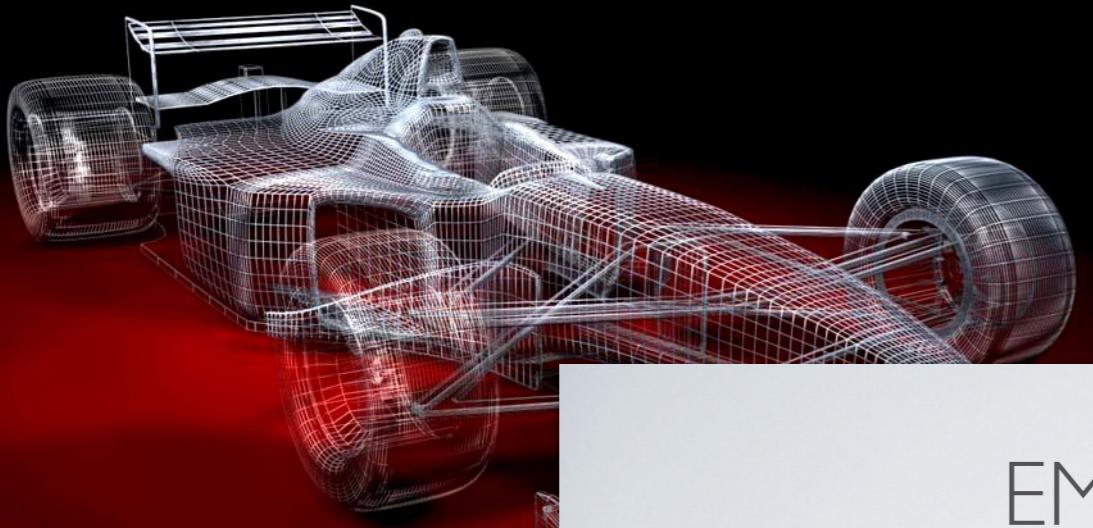
Novel algorithm

Novel approach/methodology

Problem formalization

DIFFERENT CONTRIBUTIONS

THEORETICAL



PRACTICAL

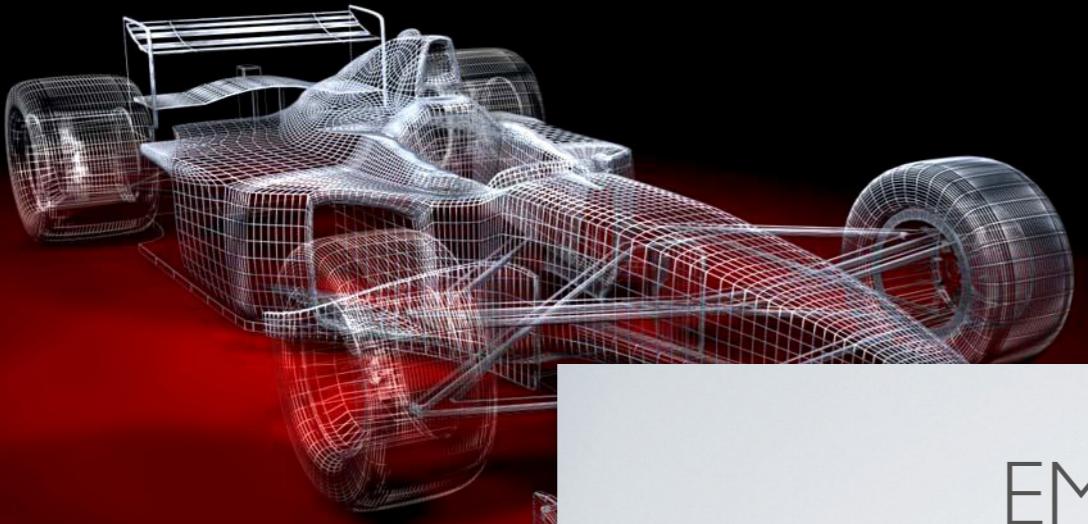


EMPIRICAL



DIFFERENT CONTRIBUTIONS

THEORETICAL



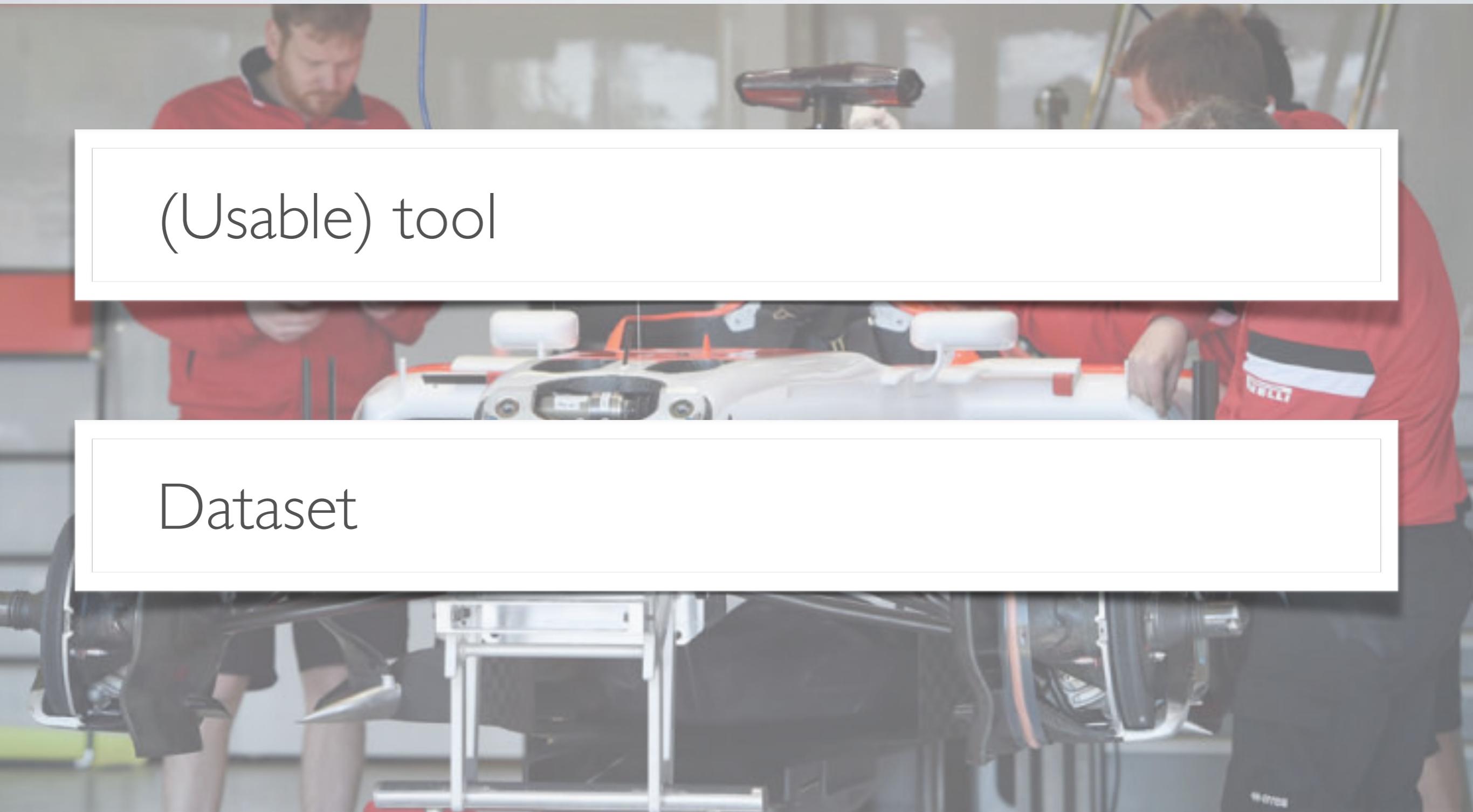
PRACTICAL



EMPIRICAL



PRACTICAL

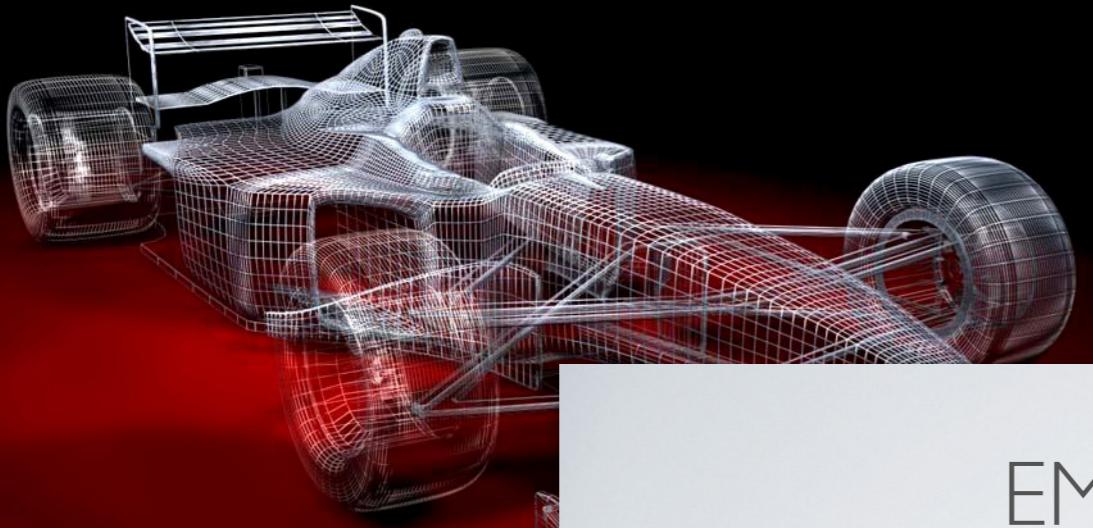
A photograph showing two men in red jackets working on a white racing kart. One man is in the foreground, focused on the front wheel area, while the other is in the background, working on the rear. The kart has orange and black accents and is equipped with a roll cage and a large front wing.

(Usable) tool

A close-up photograph of the rear wheel assembly of the racing kart. The wheel is black with a multi-spoke design, mounted on a dark frame. The background is blurred, showing parts of the kart's bodywork and engine.

DIFFERENT CONTRIBUTIONS

THEORETICAL



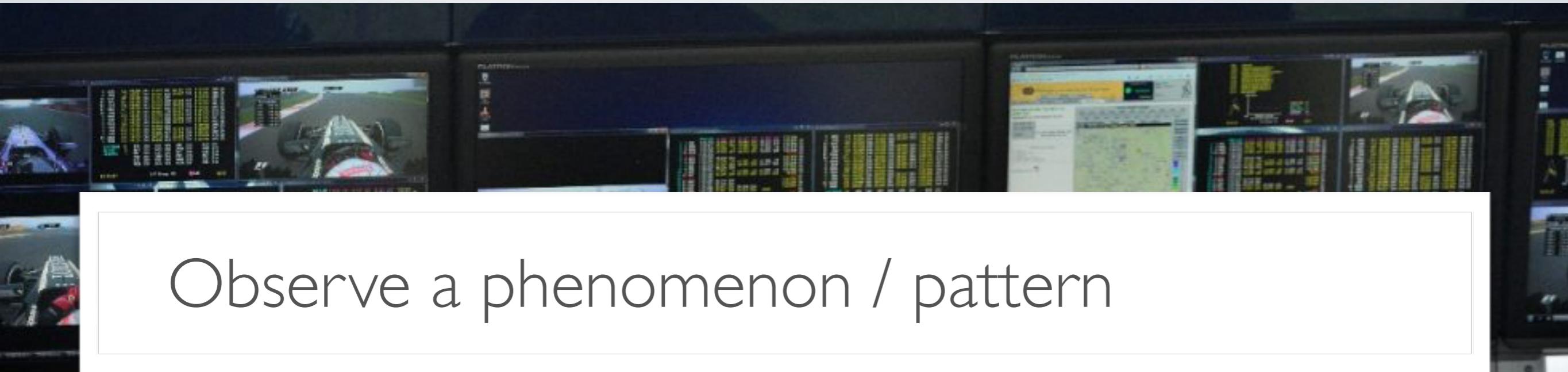
PRACTICAL



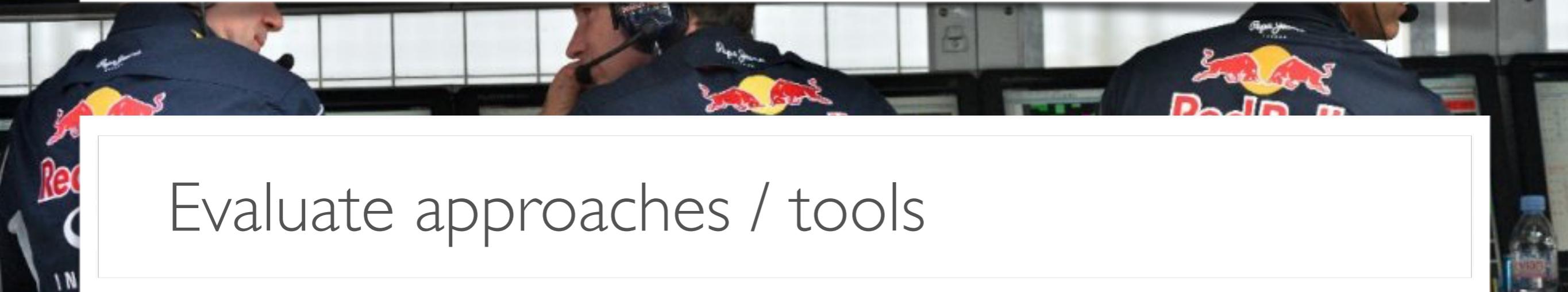
EMPIRICAL



EMPIRICAL



Observe a phenomenon / pattern



Evaluate approaches / tools



HINTS (FOR AUTHORS AND REVIEWERS)

A **good paper** may be weaker on one dimension,
but then it should be stronger on the other side



WHAT DOES IT MEAN?

- Your approach provides a small theoretical increment
- You should provide a strong empirical evidence that your approach is better than the current state of the art

CHALLENGES

- No tool available for alternative approaches
- No data
- The paper does not contain enough detail to reimplement the approach
- Asked the authors, but they don't respond

MONACO '84:A RELATIVELY MINOR
CAR AND A GREAT DRIVER...



THE RESEARCH PAPERS I LIKE

A **great paper** contributes along all dimensions

THE IDEAL PATH

1. You conceive a novel approach
2. You evaluate the approach
3. You actually implement it in a usable / downloadable tool
4. You show that the tool actually helps



Even if you don't have all these in a paper, it would be great to have them in your thesis

EMPIRICAL



EVALUATING YOUR RESEARCH



TYPICAL QUESTIONS YOU ASK

HOW ACCURATE IS IT?



HOW FAST IS IT?



IS IT ANY BETTER THAN COMPETITOR'S TOOL?



WHAT ARE WE MISSING HERE?



RECOMMENDER SYSTEMS

“A software application that provides information items estimated to be valuable for a software engineering task in a given context”

Martin P. Robillard, Robert J. Walker, Thomas Zimmermann:
Recommendation Systems for Software Engineering. IEEE Software 27(4): 80-86 (2010)

RECOMMENDER SYSTEMS

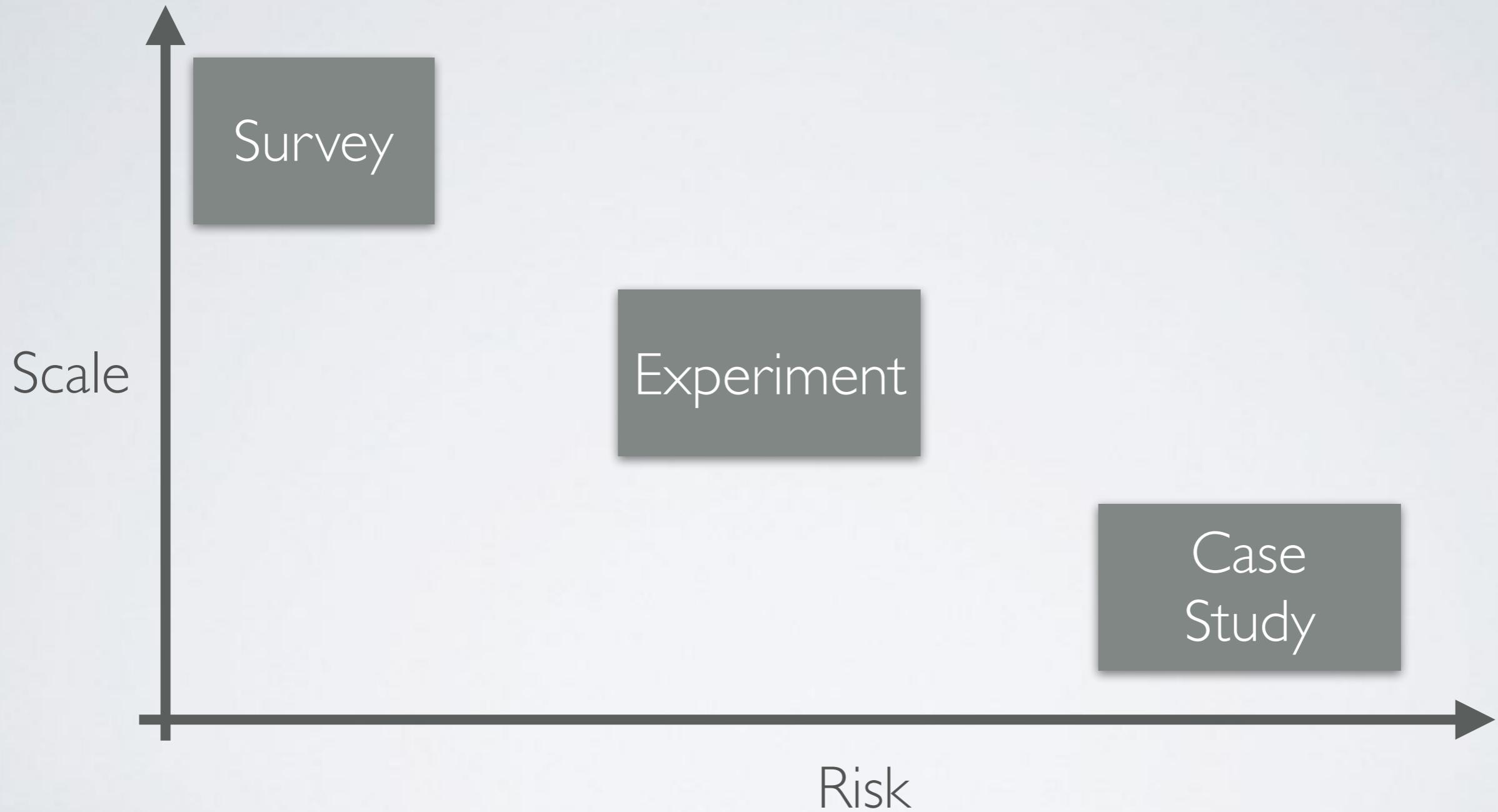
“A software application that provides information items estimated **to be valuable for a software engineering task in a given context**”

Martin P. Robillard, Robert J. Walker, Thomas Zimmermann:
Recommendation Systems for Software Engineering. IEEE Software 27(4): 80-86 (2010)

IS THE TOOL GOING
TO HELP A DEVELOPER
FOR A GIVEN TASK?

DOES IT WORK IN MY
SPECIFIC COMPANY?

DIFFERENT KINDS OF STUDIES



[Linkerman and Rombach, 2000]

EXAMPLE

Mining StackOverflow to Turn the IDE into a Self-Confident Programming Prompter

Luca Ponzanelli¹, Gabriele Bavota², Massimiliano Di Penta², Rocco Oliveto³, Michele Lanza¹

1: REVEAL @ Faculty of Informatics – University of Lugano, Switzerland

2: University of Sannio, Benevento, Italy

3: University of Molise, Pesche (IS), Italy

ABSTRACT

Developers often require knowledge beyond the one they possess, which often boils down to consulting sources of information like Application Programming Interfaces (API) documentation, forums, Q&A websites, *etc.* Knowing what to search for and how is non-trivial, and developers spend time and energy to formulate their problems as queries and to peruse and process the results.

We propose a novel approach that, given a context in the IDE, automatically retrieves pertinent discussions from Stack Overflow, evaluates their relevance, and, if a given confidence threshold is surpassed, notifies the developer about the available help. We have implemented our approach in PROMPTER, an Eclipse plug-in. PROMPTER has been evaluated through two studies. The first was aimed at evaluating the devised ranking model, while the second was conducted to evaluate the usefulness of PROMPTER.

problems, the main one being the absence of automation: Every time developers need to look for information, they interrupt their work flow, leave the IDE, and use a Web browser to perform and refine searches, and assess the results. Finally, they transfer the obtained knowledge to the problem context in the IDE. The information is retrieved from different sources, such as forums, mailing lists [2], blogs, Q&A websites, bug trackers [1], *etc.* A prominent example is Stack Overflow, popular among developers as a venue for sharing programming knowledge. Stack Overflow is vast: In 2010 it already had 300k users, and millions of questions, answers, and comments [23]. This makes finding the right piece of information cumbersome and challenging.

Recommender systems [33] represent a possible solution to this problem. A recommender system gathers and analyzes data, identifies useful artifacts, and suggests them to the developer. Seminal

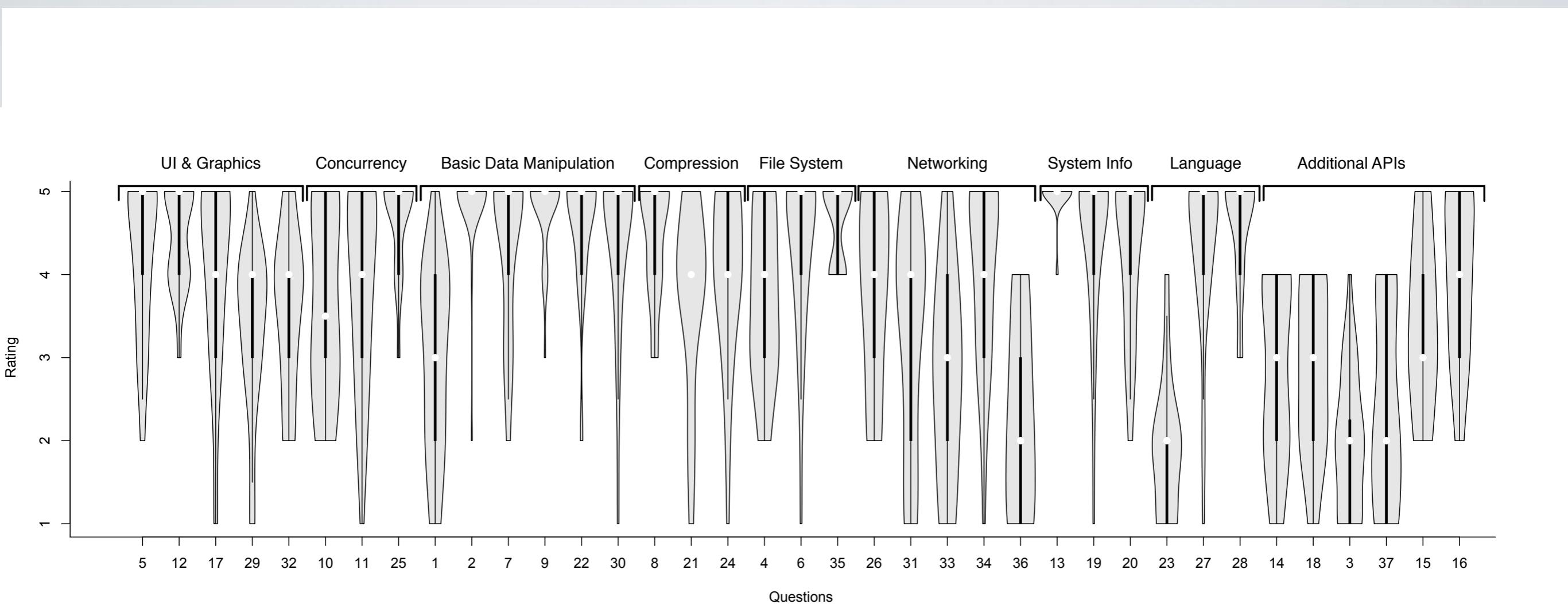
Luca Ponzanelli, Gabriele Bavota, Massimiliano Di Penta, Rocco Oliveto, Michele Lanza:
Mining StackOverflow to turn the IDE into a self-confident programming prompter. MSR 2014:
102-111

TOOL (PROMPTER)

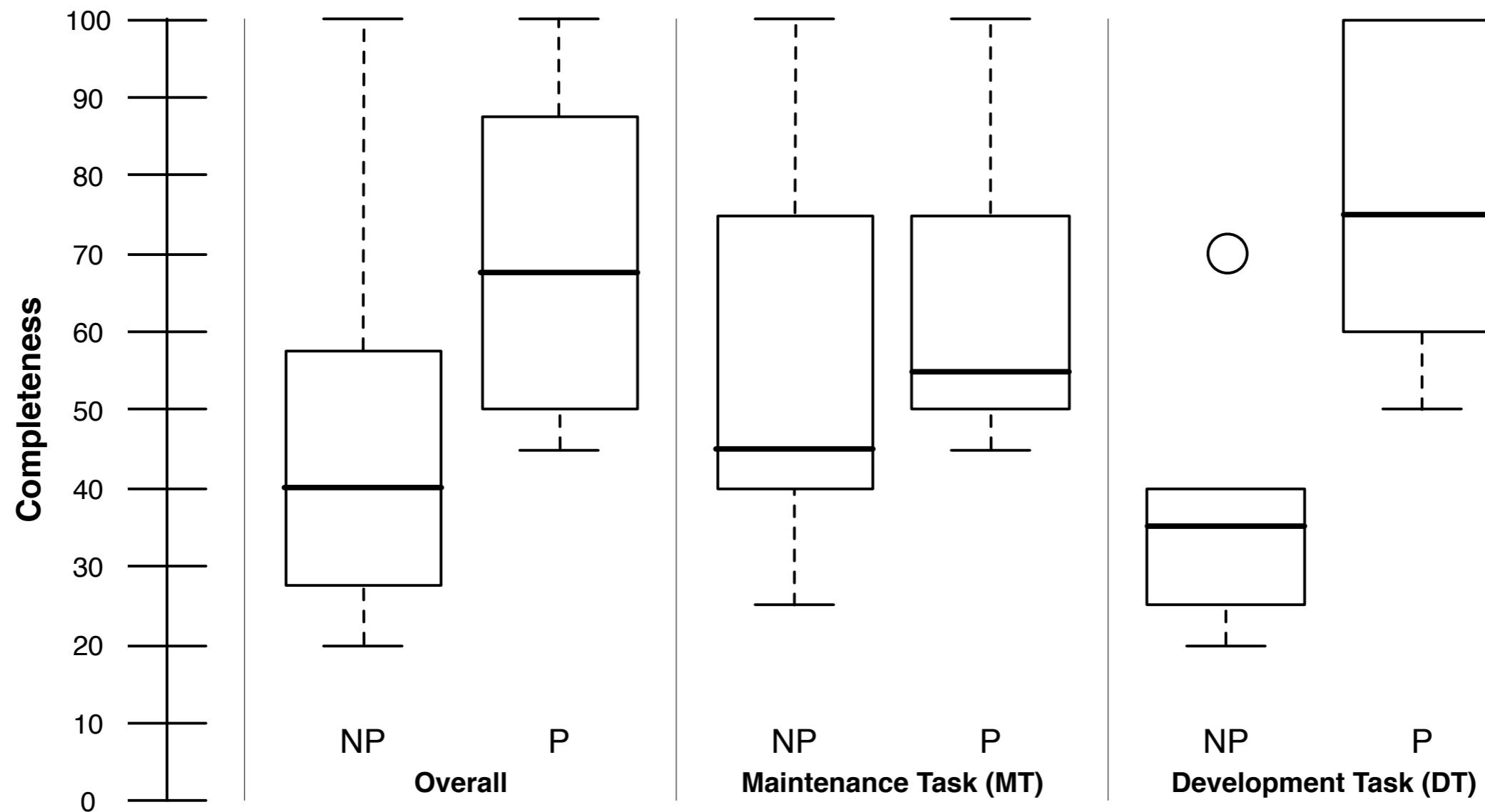
The screenshot shows a developer's environment with several windows open:

- Code Editor:** A Java code snippet for unzipping files. The code uses `ZipInputStream` to read from a ZIP file and `FileOutputStream` to write to a folder. It includes annotations for parameters and a note about creating the output directory if it doesn't exist.
- Notification Center:** A window titled "Notification Center" showing two notifications. The first notification is for "Java ZIP - how to unzip folder?" with a progress bar at 66% and a timestamp of 6/9/2013 16:04. The second notification is for "How to add a progress bar?" with a progress bar at 61% and the same timestamp. Both notifications have a red circle with the number "1" in the top right corner.
- Stack Overflow Document:** A window titled "Stack Overflow Document" containing a question about Java ZIP. The question asks for sample code to unzip a folder from ZIP into a desired directory. It includes a rating section with five stars and a "Send" button, and a footer with a "2" icon and tags "java" and "zip".

ARE RECOMMENDATIONS RELEVANT?



DOES IT HELP?



MANAGING YOUR PHD



THE ADVISOR



DON'T EXPECT FROM YOUR ADVISOR

- Dealing with low-level details
- Carrying out your research
- Deciding what is good for your future

EXPECT FROM YOUR ADVISOR

- Regular feedback
- Hints, also from a strategical point of view
- Some seminal ideas...
- Decreasing contributions along your PhD

NOT GOING TO HAPPEN



YOU ARE THE DRIVER!



COLLABORATIONS



COLLABORATIONS



Computer science can be seen as a very theoretical discipline where individual work dominates...

GREAT ENGINEERING IS A RESULT OF TEAM WORK



REASONS TO LIKE COLLABORATIONS

- Exchange of ideas, working habits, culture
- Opportunity to shine
- Network building
- Different competencies combined together
- Joining the force for human-intensive work

SUGGESTIONS

Be flexible: different people have different working habits

Be open to others' ideas and opinions

WARNING

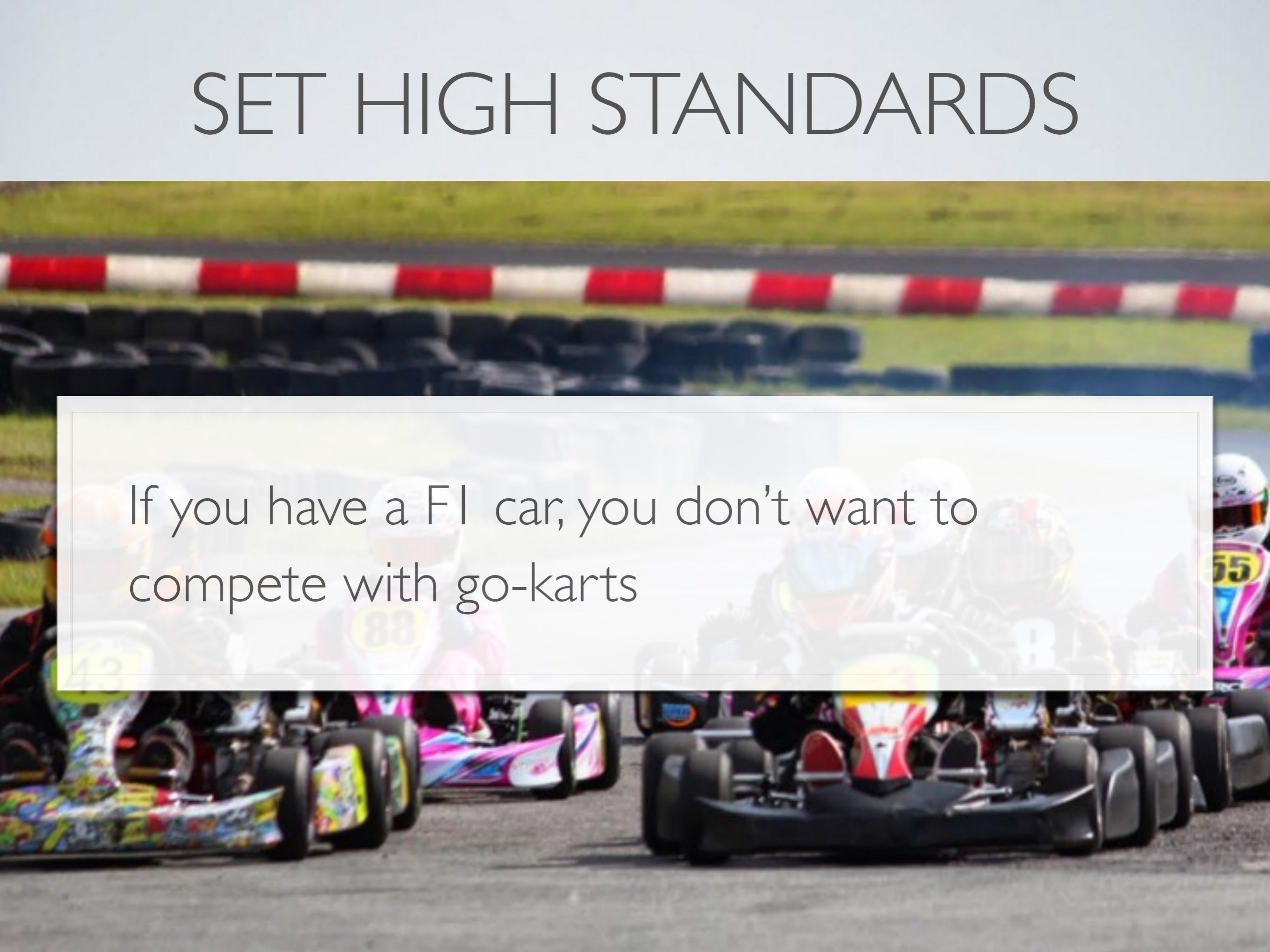


- Make sure your contribution is clear
- You need to be the leader in some of your works

SET HIGH STANDARDS

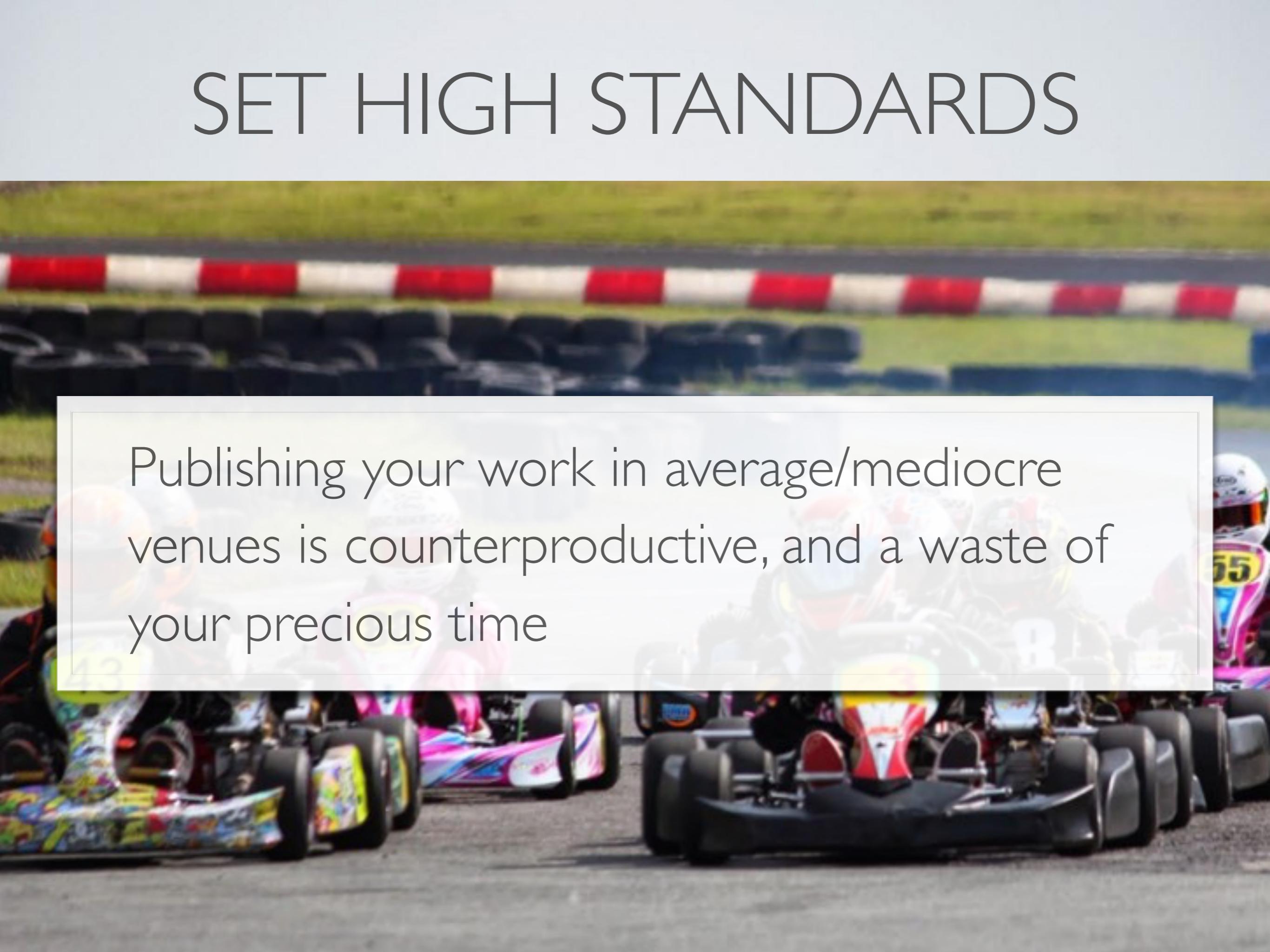


SET HIGH STANDARDS



If you have a F1 car, you don't want to compete with go-karts

SET HIGH STANDARDS



Publishing your work in average/mediocre venues is counterproductive, and a waste of your precious time

WHAT ARE THE BEST VENUES?

Ask yourself since day zero

Always target top venues at first



DIFFERENT OPTIONS

- **Top mainstream** journal and conferences
- **Specialized** good conferences
- **Workshops** are good for discussing ideas, better than mediocre conferences

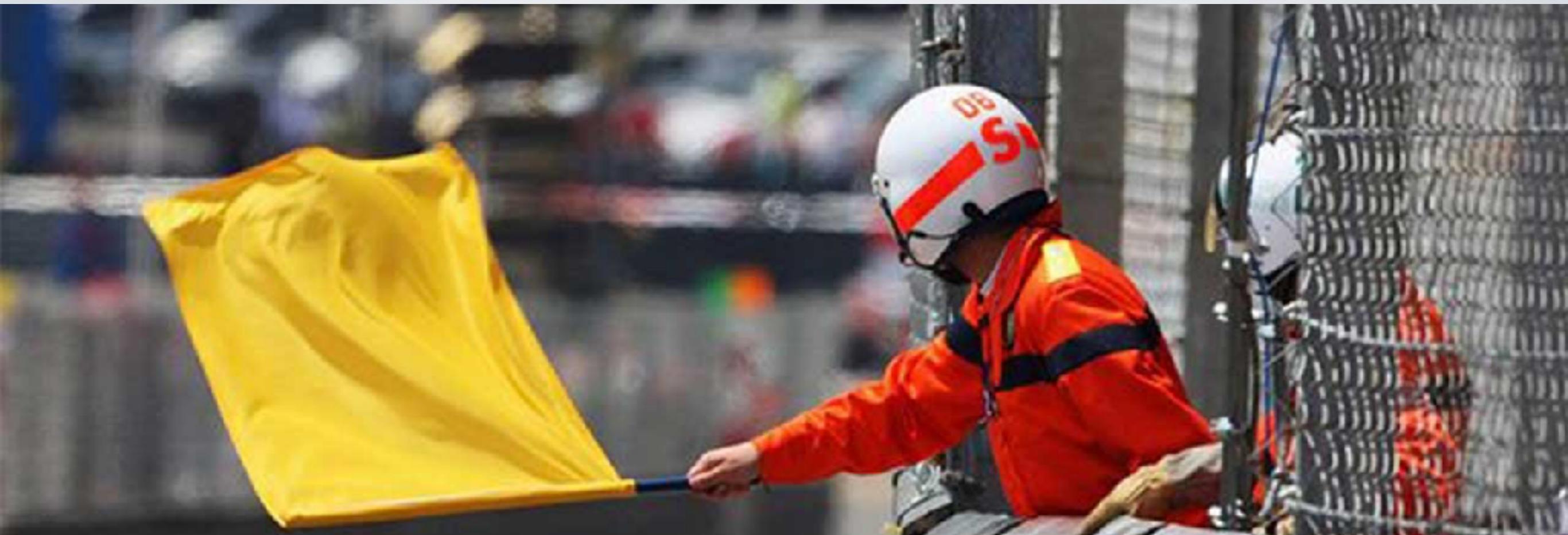
CONFERENCES VS JOURNALS

- Timeliness
- Visibility(*)
- You have a deadline

- Universally recognized
- Plenty of space
- Multiple revision rounds

(*) Now also true for journals (journal-first option)

BE AWARE



- Different institutions/countries → different evaluation standards
- Top conference vs. high impact factor journals
- A mix of both doesn't hurt

CONTINGENCY PLANS



RESEARCH VS PRODUCTION



- Research is not straight-forward
- You could spend months on a solution that reveals to be wrong or inefficient



PETERSON

FIRST NATIONAL CITY
TRAVELERS CHECKS

3

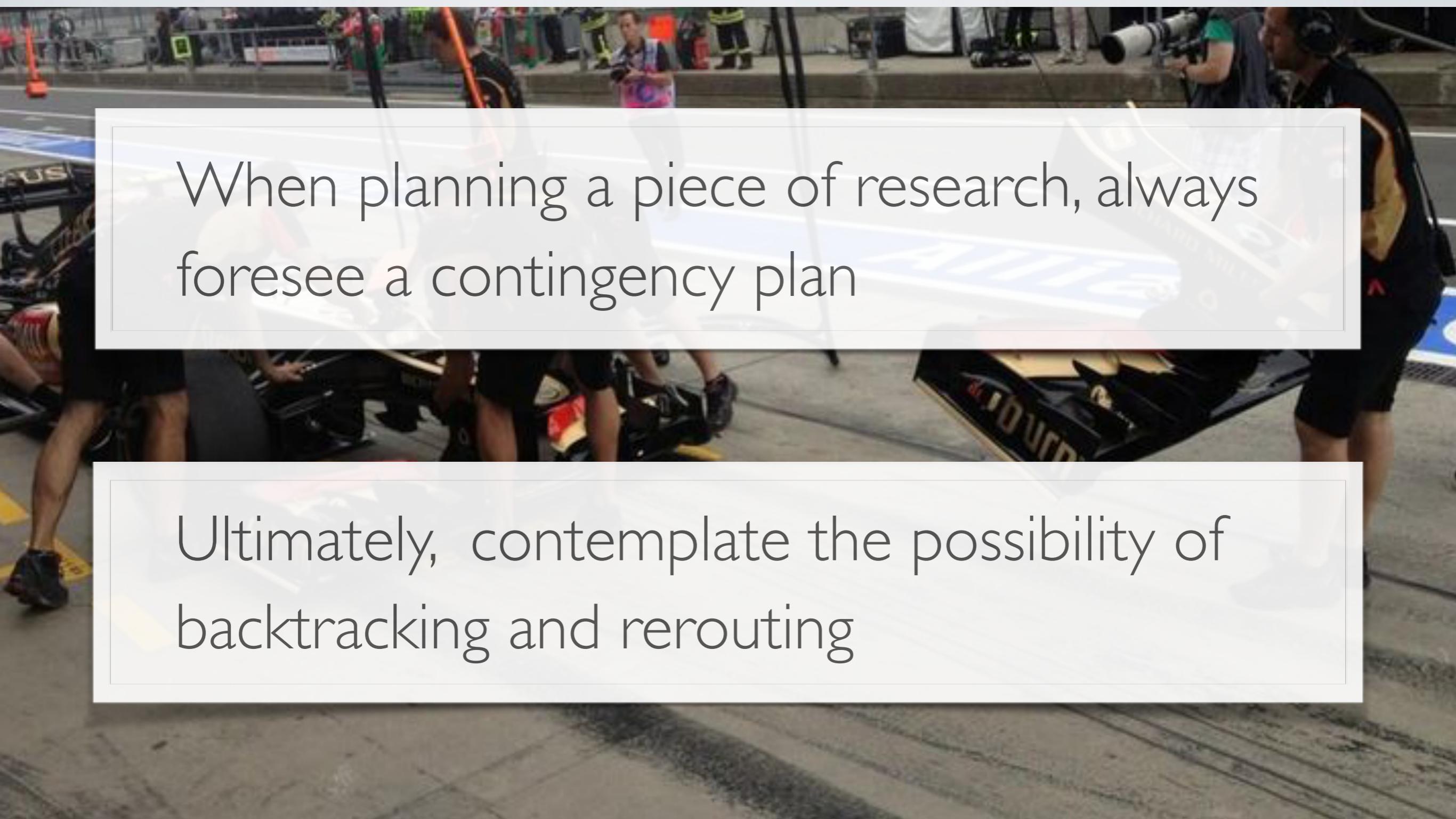
elf

elf

DON'T BLAME YOURSELF NOR
YOUR ADVISOR!



HINTS



When planning a piece of research, always foresee a contingency plan

Ultimately, contemplate the possibility of backtracking and rerouting

NEGATIVE RESULTS

Negative results are worthwhile of being published

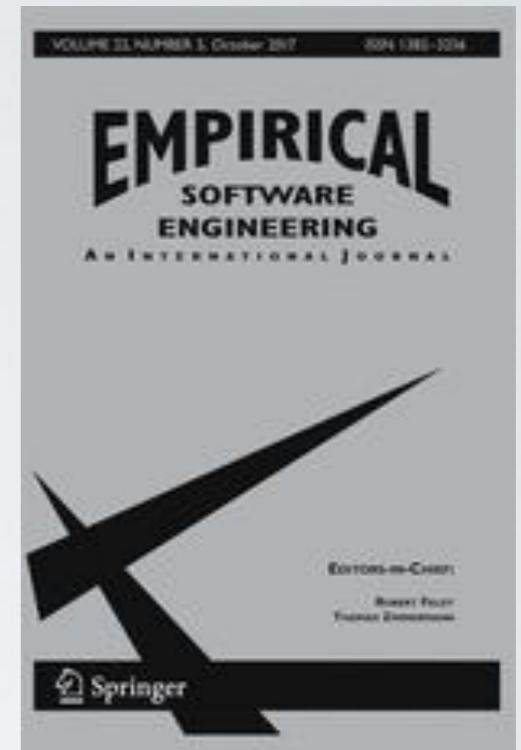
WHAT NEGATIVE RESULTS COULD TELL US

A very successful technique completely fail in
a specific context

An approach is not ready for real-world
application

SOME EXAMPLES

Special Section of the EMSE Journal,
October 2017



<https://link.springer.com/journal/10664/22/5/page/1>

DOES SENTIMENT ANALYSIS WORK FOR SE TASKS?

On negative results when using sentiment analysis tools for software engineering research

Robbert Jongeling¹ · Proshanta Sarkar² ·
Subhajit Datta³ · Alexander Serebrenik¹ 

Published online: 10 January 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract Recent years have seen an increasing attention to social aspects of software engineering, including studies of emotions and sentiments experienced and expressed by the software developers. Most of these studies reuse existing sentiment analysis tools such as SENTISTRENGTH and NLTK. However, these tools have been trained on product reviews and movie reviews and, therefore, their results might not be applicable in the software engineering domain. In this paper we study whether the sentiment analysis tools agree with the sentiment recognized by human evaluators (as reported in an earlier study) as well as with each other. Furthermore, we evaluate the impact of the choice of a sentiment analysis tool on software engineering studies by conducting a simple study of differences in issue resolution times for positive, negative and neutral texts. We repeat the study for seven datasets (issue trackers and STACK OVERFLOW questions) and different sentiment analysis tools and observe that the disagreement between the tools can lead to diverging conclusions. Finally, we perform two replications of previously published studies and observe that the results of those studies cannot be confirmed when a different sentiment analysis tool is used.

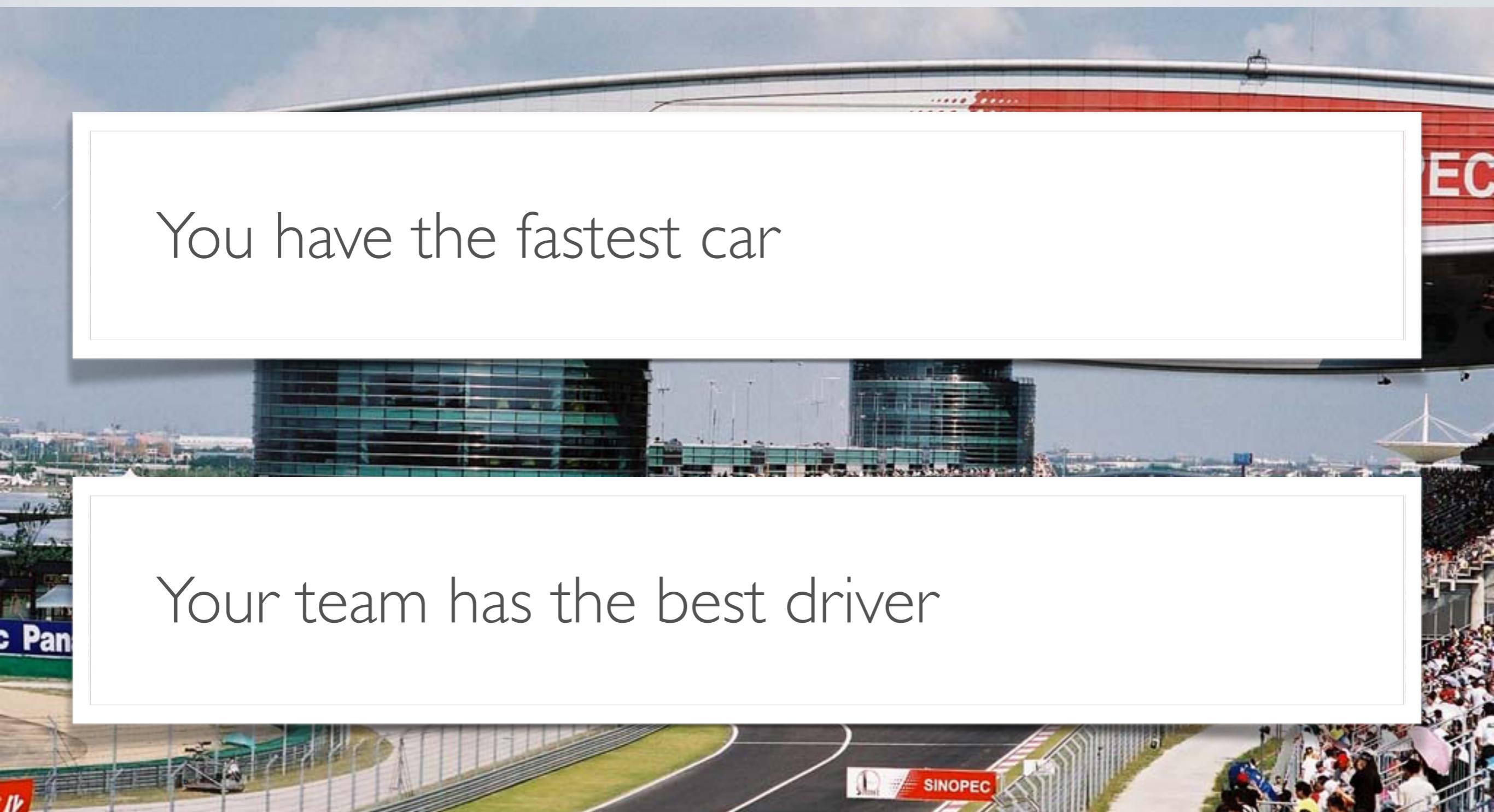
SANER 2018 RENE TRACK



SANER 2018 <http://saner.unimol.it/>

REproducibility Studies and NEgative Results (RENE) Track

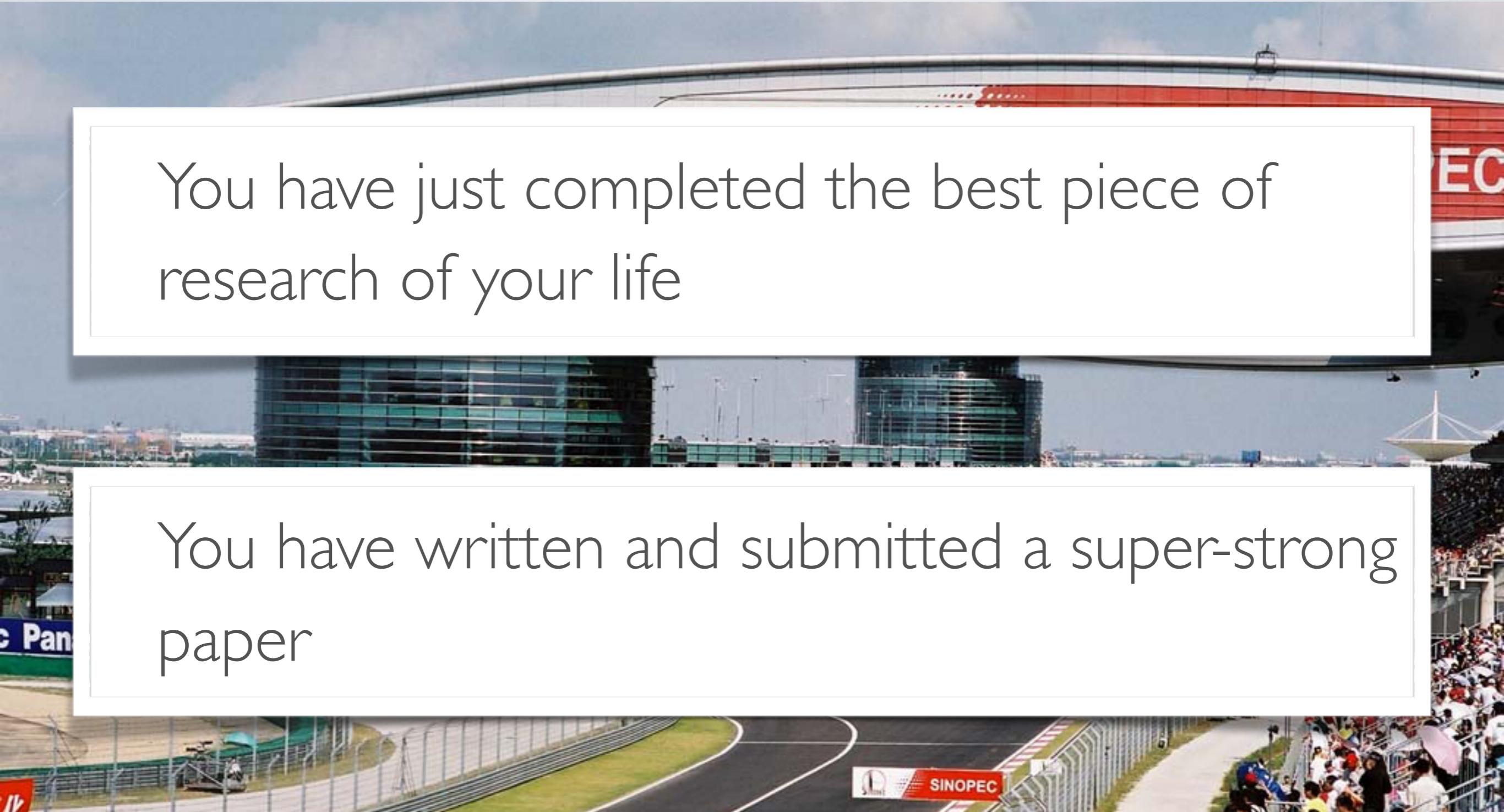
CONFERENCE SUBMISSIONS ARE LIKE A RACE



You have the fastest car

Your team has the best driver

CONFERENCE SUBMISSIONS ARE LIKE A RACE



You have just completed the best piece of research of your life

You have written and submitted a super-strong paper

REJECT



LESSONS

If your work is not good, chances to be accepted in a good venue are fairly limited

If your work is good, the unpredictable can still happen....

WHY?

- Mild vs super-picky reviewers
- Non-expert (but still qualified) reviewers
- Having (or not) a champion or a strong detractor

HINTS

- Never complain or even ask to reconsider
- Reviews contain mainly good suggestions, be critical with yourself and benefit from them
- Ok, not all suggestions might be good...

MORE IMPORTANT

Take a rejection as an opportunity to improve
your work, not as a suggestion to downgrade

YOU MAY LOOSE A RACE...



But what matters is winning the championship!

PHD ACCOMPLISHED!



WHAT'S NEXT?



POST-DOCVS FACULTY POSITION



It really depends on what's your target and life decision

MAIN PURPOSE OF POST-DOC



It's like Drag Reduction System (DRS) in Formula 1

MAIN PURPOSE OF POST-DOC



- Gives you the boost to compete for a very prestigious position
- You have very limited time, don't waste it

HOW YOUR WORK WILL
CHANGE AFTER YOUR PHD

FEELING RELIEVED?

- I'm done!
- No more overnight work, crazy deadlines...
- Now I don't have a boss anymore!

NEW RESPONSIBILITIES

A photograph showing the lower legs and feet of a person submerged in clear, light blue water. The person is wearing dark swim trunks. The background consists of rippling water.

Teaching

Service

Fund raising

TEACHING



TEACHING

- Along with research, your primary job if you stay in academia
- Teaching priorities vary depending on the institutions

HINTS

- Don't oversell yourself / overcommit
- Have fun and let your students have fun
- Connect your teaching with your research
- Don't teach what students can find on the Internet anyway

SERVICE



SERVICE TO YOUR INSTITUTION

- I agree this is by far the most boring part, but...
- Without it, excellence in education and research won't be possible

SERVICE TO YOUR COMMUNITY

Reviewing

Conference organization

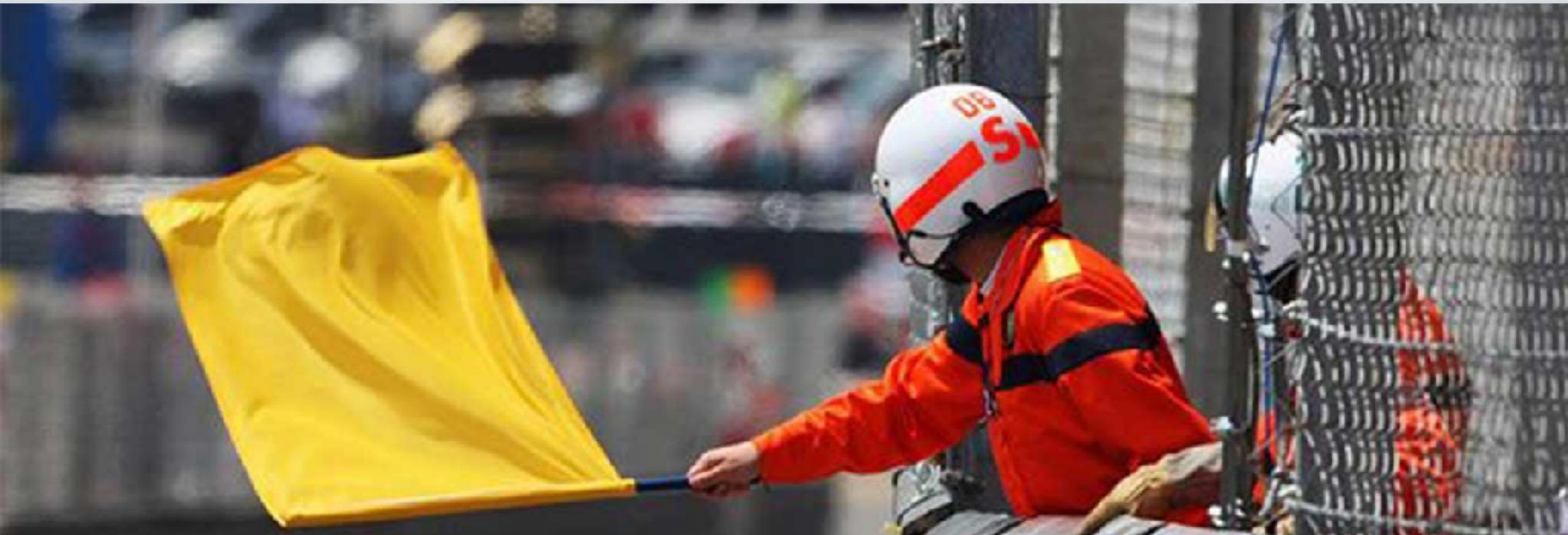
Other

REVIEWING



- I admit I do a lot
- I care about my reviewing duties as much as I care about my research

PROBLEMS



- Volume of submissions continuously increasing
- Not everybody available to reviewing

HARD TO GET (GOOD) REVIEWERS

- Little credit, especially for journal reviewing
- Too much work in a very short time
- Conference reviewing might involve expensive PC meetings

CONFERENCES AND
JOURNALS WON'T EXIST
WITHOUT REVIEWERS!

SO....

- If you write papers, you should also review
- Don't overcommit and feel free to say no, but not always
- Be timely / provide updates about your timeline

WRITING GOOD REVIEWS

- Find reasons for acceptance, not for rejection
- **Don't write reviews you wouldn't like to receive**
- Explain why the paper should or should not be accepted
- Provide feedback to improve the paper

FUND RAISING



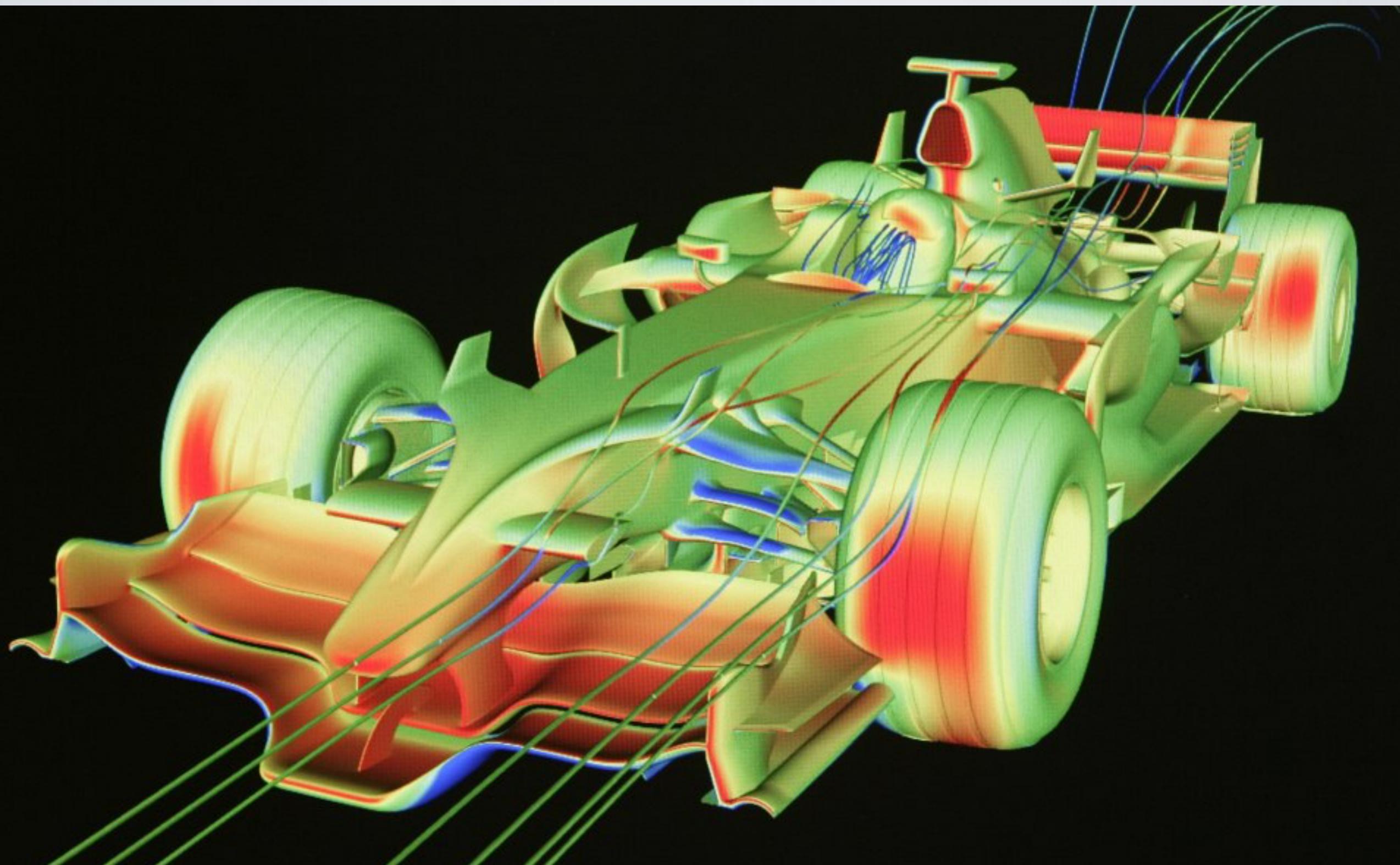
FUND RAISING

- Way too many models
- You need to get familiar with the models of the country/area where you (will) work

HINTS

- Fund raising functional to your research, not the other way around
- However, sometimes you need to fit your topics to grant calls

WHAT ABOUT RESEARCH?



WHAT ABOUT RESEARCH?

- Time to show your ability to develop your own path
- However, attacking a completely new topic takes time
- Make sure your research plan fits with your tenure timeline

CHANGING ROLE



CHANGING ROLE

You will get busier and busier pretty soon!

Bad news: less time to do things you really like

MY PERSONAL SUGGESTION

Don't "abstract" yourself completely or you will loose contact with the reality

ADVISING YOUR STUDENTS

- Always remember when you were a student
- Learn from your mistakes and from your advisor's mistakes

STILL SOMETIME LEFT FOR ME?



YES YOU NEED TO REFUEL!



MANAGING LIFE

- Allocate time for your personal life
- Alternate periods in which you work under pressure with relatively calm period
- That's when you often have the best ideas!

FINAL REMARKS



Contributions along different dimensions

Things can go wrong, keep it into account

Service is as important as research and teaching

Don't just become a manager

Questions?

dipenta@unisannio.it, [@maxdipenta](https://twitter.com/@maxdipenta)