## Planning:

- Problem refinement
- Collect data → Clean & preprocess data
- Feature & model selection
- Train & eval results → Deploy

## Data:

What prediction problem would you like to solve?

- Classification? Regression?

What kind of data is needed to solve?

- Time series
- Text data
- Tabular
- Img data
- Sound

Open Source Dataset: Kraggle, etc ...

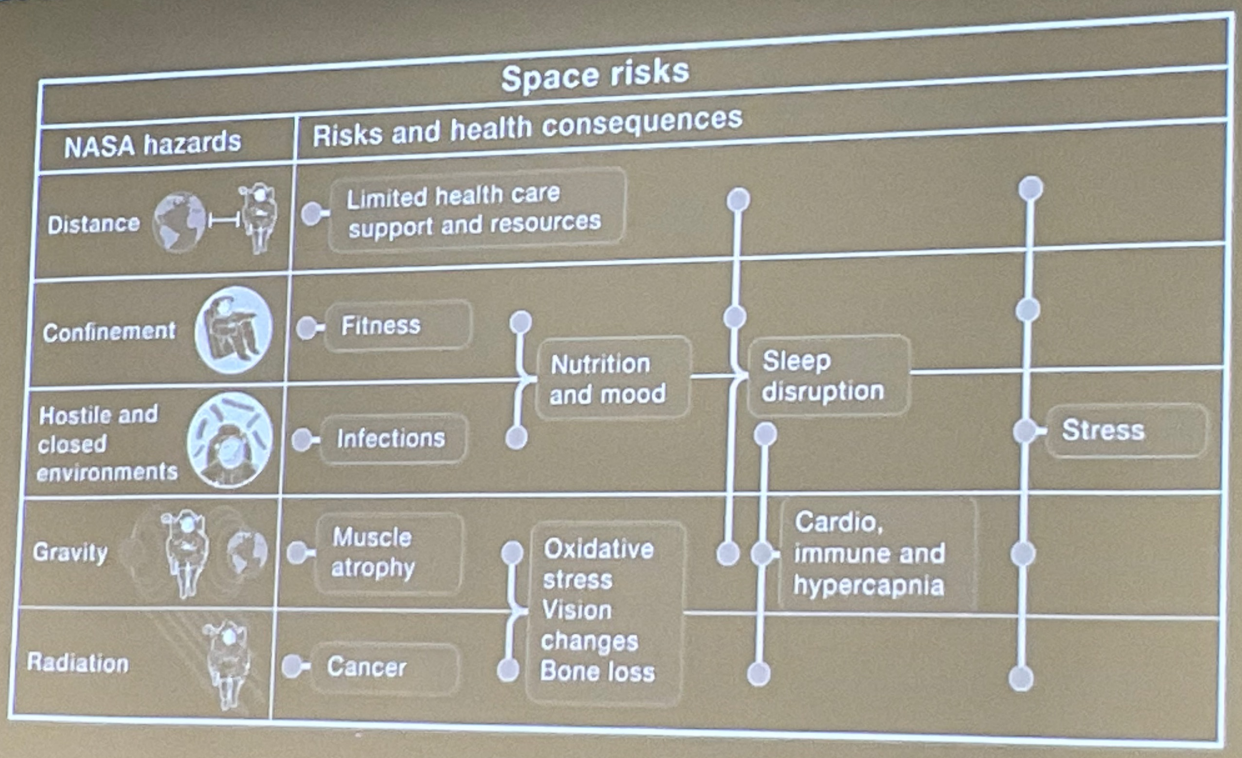
## Tooling & Infrastructure:

Data → Development → Deployment | Github

(???) (Pytorch → PT Lighting) (Stream Lit)

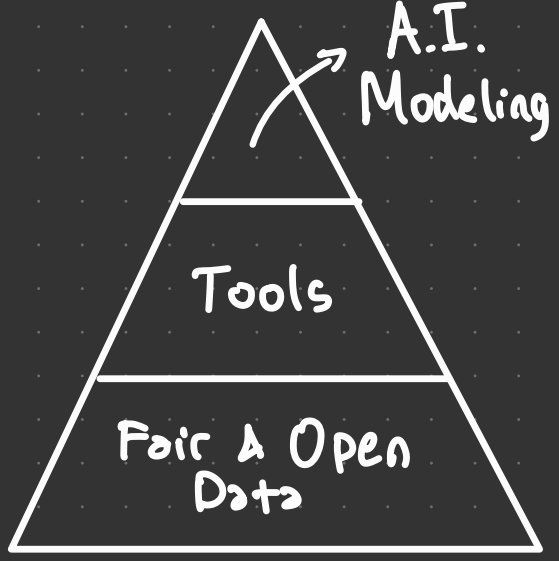
# CHALLENGE UNVEILING: Astronaut Health!

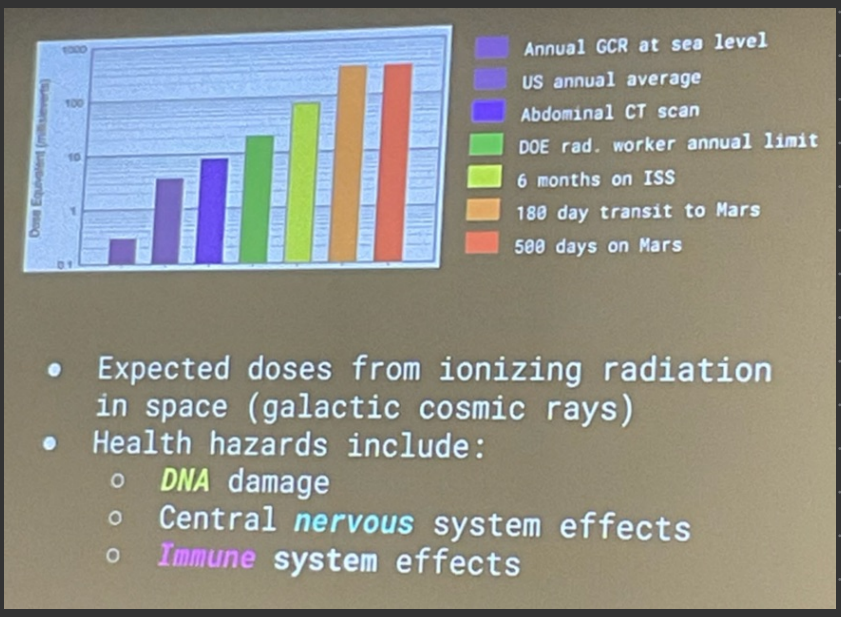
**Exposure to spaceflight is associated with a set of health impacts due to 5 key "stressors:"**




Afshinnekoo et al., Cell 2020

**Goal:** Leverage ML to understand Animal Bio
Micro Bio, Dev - Reproduce - Evol Bio, Mole Bio



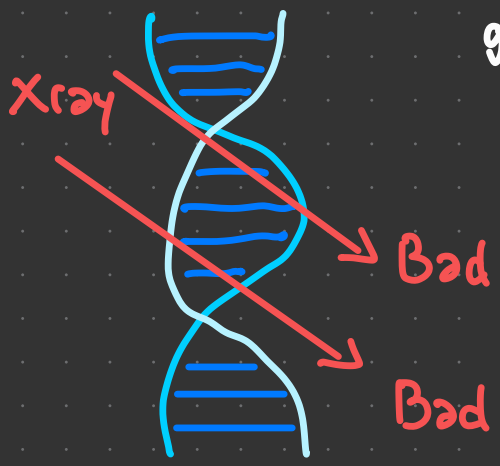




- Expected doses from ionizing radiation in space (galactic cosmic rays)
- Health hazards include:
  - *DNA* damage
  - Central *nervous* system effects
  - *Immune* system effects

Using tools like GeneLabs

# Radiation:

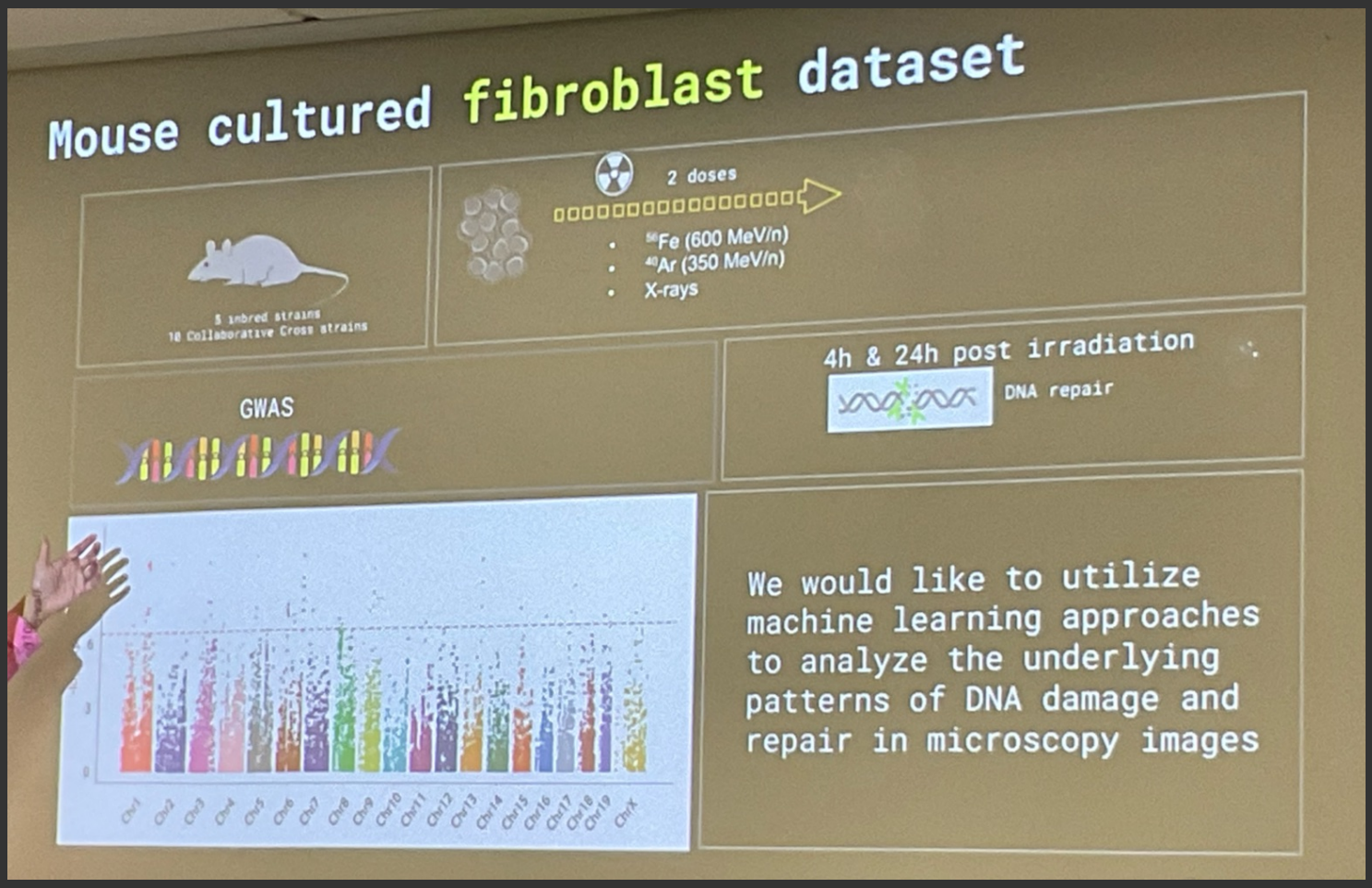
- Composed of solar particle events (SPEs)
  galactic cosmic rays (GCRs)



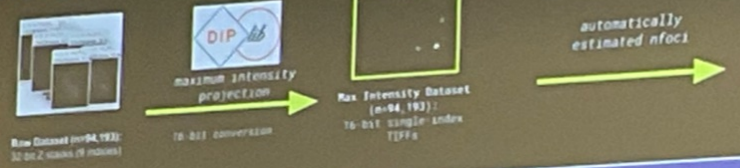

87% protons
4% $^{4}He$
~1% $^{56}Fe$

Visible through fluorescent imging of DNA markers

- Study on mice ⇝ Cheap, similar genetically



# Mouse fibroblast DNA damage benchmark dataset




| Label Types | Labels | Total images |
|---|---|---|
| Number of foci (nfoci) | 0-20 | 93,488 |
| Radiation Type | $^{56}$Fe or X-ray | |
| Radiation Dose | Low and high dose | |
| Imaging Time Post-exposure | 4, 24, 48 hours | |

Registry of Open Data on AWS

aws

BPS Microscopy Benchmark Dataset