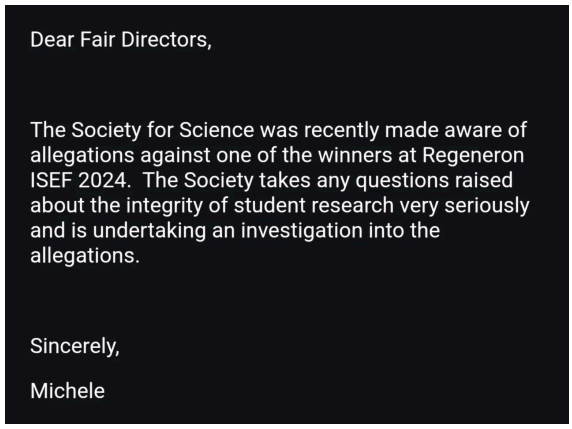
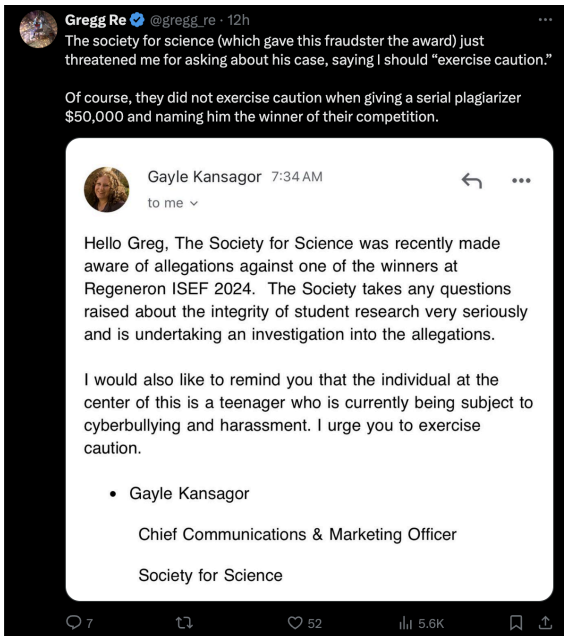
Updates about the situation are shared on Twitter. As of May 21:



Society for Science has not yet made a public response, however, some journalists have reached out to them for private comment. Exercise caution.



Dear Society for Science, Regeneron, and other Regeneron ISEF Affiliates:

According to the rule book, "Scientific fraud and misconduct are not condoned at any level of research or competition. This includes plagiarism, forgery, use or presentation of other researcher's work as one's own and data fabrication. Fraudulent projects will fail to qualify for competition in affiliated fairs and Regeneron ISEF. Society for Science and the Public reserves the right to revoke recognition of a project subsequently found to have been fraudulent."

This year, among the top winners of the fair, "**ENEV081 - Efficiently Discovering Plastic-Degrading Microbes**" stands out due to (1) direct evidence of fraud and falsified claims; (2) blatant plagiarism and citation fraud; (3) potential scientific inaccuracies. According to Gregg Re, lawyer at the Daily Wire, the position maintained is that these issues were "innocent citation errors."

This project was awarded with the Regeneron Young Scientist Award, a $50,000 scholarship, and First Place, a $5,000 cash award. Because research fraud is a major concern in academia, it is very important for ISEF to at least acknowledge this issue.

Regeneron and Society for Science have shared this project in Business Insider, among other articles:

- https://youtu.be/JIRb2SGqC-U?si=DTsvnq8O3crqFZAY

- https://www.snexplores.org/article/2024-regeneron-isef-winners-bioelectronics-genetics-math
- https://www.prnewswire.com/in/news-releases/more-than-9-million-awarded-to-high-school-scientists-and-engineers-at-the-regeneron-international-science-and-engineering-fair-2024-302149316.html
- https://www.thebrighterside.news/post/16-year-old-s-revolutionary-discovery-could-make-medical-implants-safer-and-more-effective
- https://markets.businessinsider.com/news/stocks/more-than-9-million-awarded-to-high-school-scientists-and-engineers-at-the-regeneron-international-science-and-engineering-fair-2024-1033400511

All of the instances of fraud that the community has identified came from these publicly available documents from the ENEV081 project:

1. Winning press release: https://www.snexplores.org/article/2024-regeneron-isef-winners-bioelectronics-genetics-math
2. Saved archive of ProjectBoard so that changes cannot be made: https://archive.ph/chnwC
3. Project presentation, video, and more: https://projectboard.world/isef/project/enev081-efficiently-discovering-plastic-degrading-microbes
4. San Diego ISEF-affiliated fair research paper: https://gsdsef.zfairs.com/api/FileApi/View/d8332a61-ca8b-4056-8598-14664b48c66d/4251d9b4-2af0-4844-acd8-efb0ebed643a/00000000-0000-0000-0000-000000000000/4163e4e2-f47e-4c42-8605-f60ea9cb4aab

We, the writers of this letter, strongly believe in Regeneron ISEF judges, and believe they are extremely qualified. Yet, due to intentional manipulations by this finalist, as we outline in this letter, inaccuracies and falsifications can slip through.

We encourage Society for Science and Regeneron to investigate this further, establish a method to handle future cases of research fraud, and take corrective action regarding the ENEV081 project being awarded despite engaging in research fraud.

Because we do not want to target students — and despite this particular researcher clearly engaging in fraud — we will not be including any reference to the researcher's name in this letter. This is because we do not want to cause long-term reputational damage to this student, who has likely still worked hard — despite stealing figures, falsifying data, etc. in their winning Regeneron ISEF project.

In a world where research fraud is increasing rapidly (notable events these last two years in particular, with notable scientists being called out), this needs to be a priority. We also note that the finalist engaged in many of these known fraudulent behaviors:

1. Harvard's Francesca Gino (inconsistent data, faked numbers): https://www.vox.com/future-perfect/24140581/francesca-gino-research-fraud-dishonesty-meta-science
2. Stanford President Resigns (results taken from other papers, image manipulation): https://www.npr.org/2023/07/19/1188828810/stanford-university-president-resigns
3. Dana-Farber Cancer Institute Scandal (manipulating figures): https://www.nbcnews.com/science/science-news/cancer-institute-dana-farber-retracts-studies-errors-rcna143922

It is not ideal for Regeneron, Society for Science, and its affiliates to reward work that has elements of research fraud in it. This fosters a harmful culture for high school level researchers, and unfairly rewards those who have cheated.

Due to the stresses of competition, students may resort to such unfair methods, putting others at a disadvantage, and harming the scientific research process. This must be extinguished at the start of their careers (Regeneron ISEF), so that these researchers do not engage in careers of publishing falsified data — which can impact real people in the future.

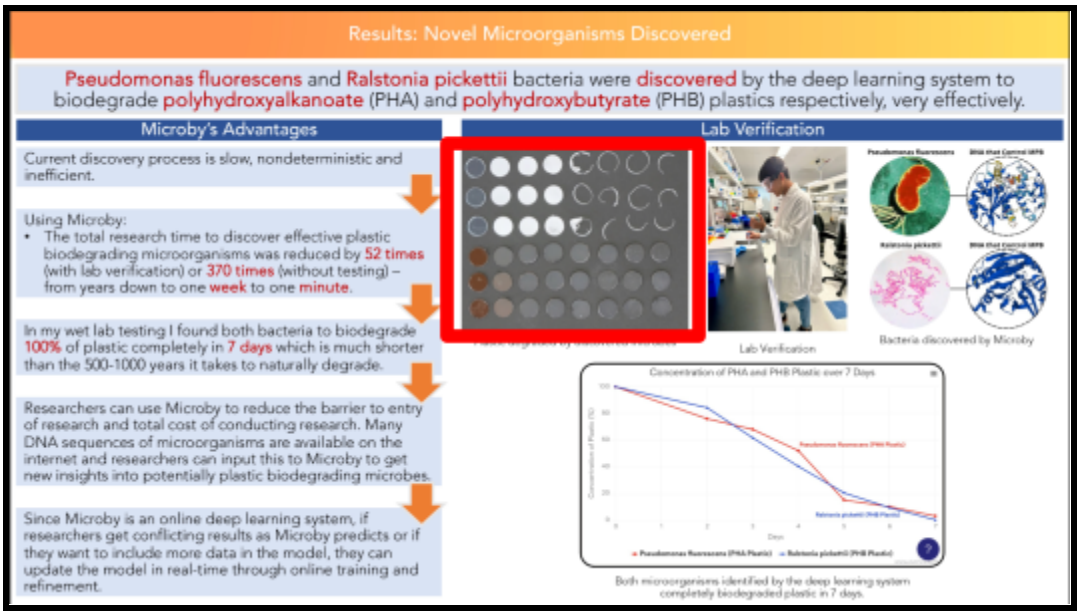
Signed,
ISEF 2024 and future competitors
ISEF & Society for Science alums
Concerned researchers

To all who wish to express their opinions/further evidence, please simply share those directly to Society for Science, Regeneron, and organizers of ISEF. These contacts can be found at the bottom of the document.

Main Issues: Direct Evidence of Falsified Claims

1. Image manipulation of result

The ISEF winner student uses the following image as a key claim of the 100% plastic degraded in their presentation and more:



Slide from student's ProjectBoard.world ISEF 2024 presentation.

However, the image boxed in red above is a **falsified image** taken from online, and has had mirroring performed in the hopes that no one would notice.

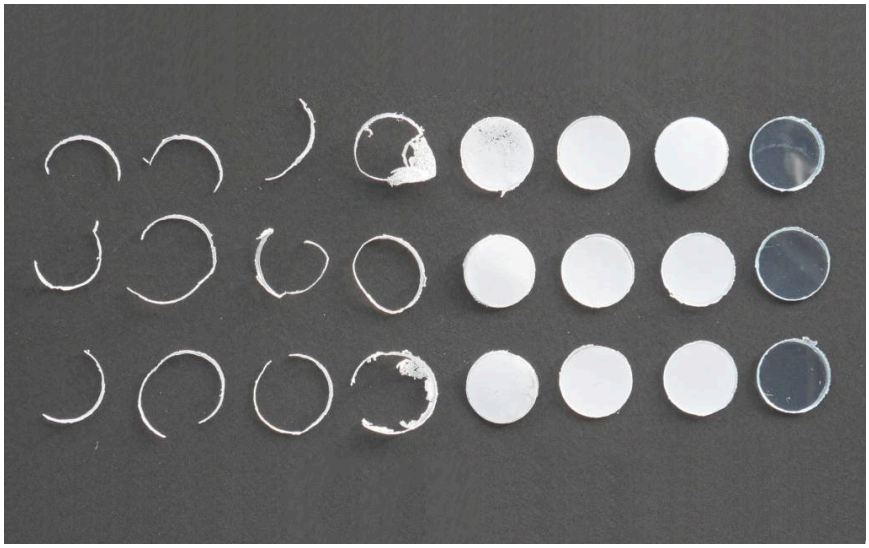
- The original image can be found here: https://www.mci.edu/en/news-filter-en/228-researchnews/4728-microorganisms-can-degrade-plastics and is from a European University Ulysseus lab testing *Ideonella sakaiensis*, a completely different organism than the one the Regeneron ISEF finalist used.

The ISEF finalist **very clearly labels** the figure as:

Plastic degraded by discovered microbes

Which is clearly false - a clear cut case of fraud. **The ISEF Finalist is taking other people's data for completely different research projects, and claiming it as their own.**

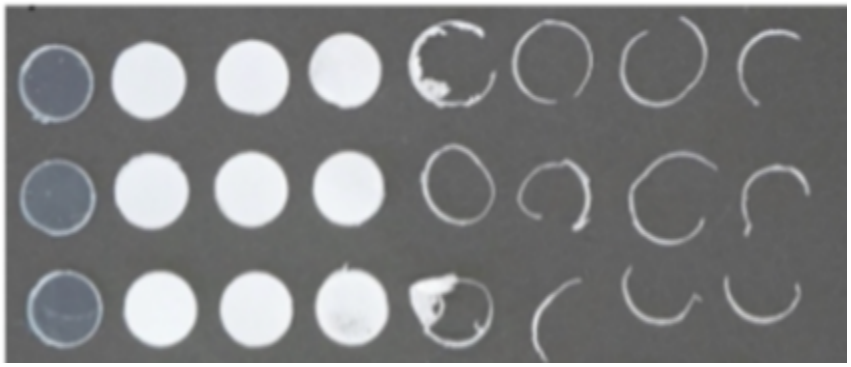
Original image from European University Ulysseus lab:



Original image from [Walter et al., 2022]

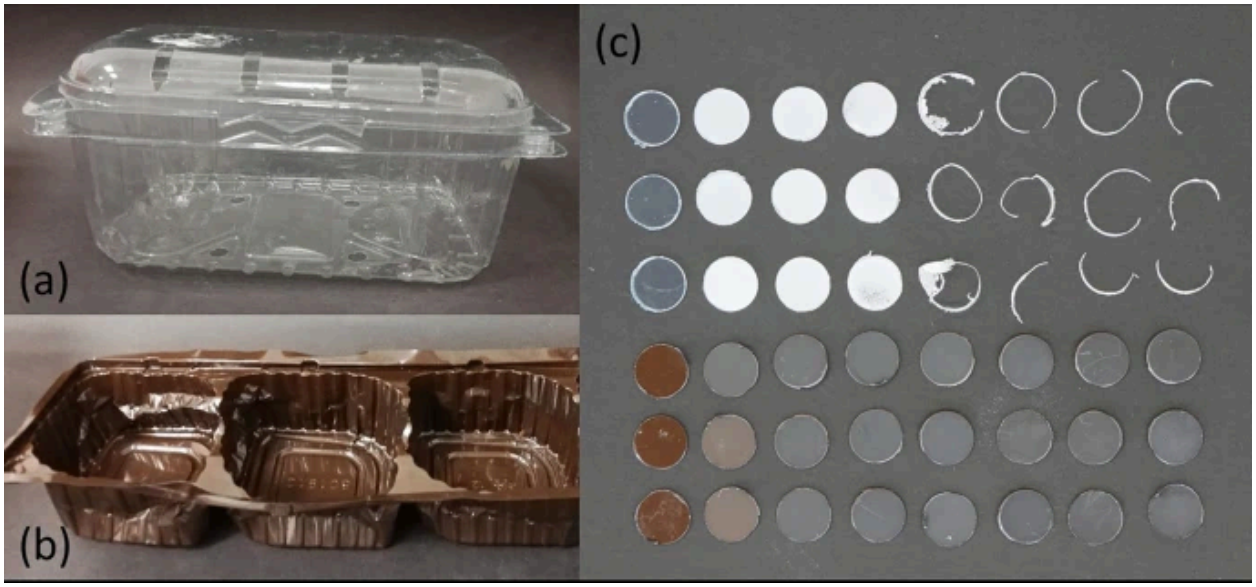
The ISEF finalist mirrored the image (which requires flipping on both axes) to generate the below image shown on their poster, presentation, etc:

Falsified, mirrored image by ISEF finalist:



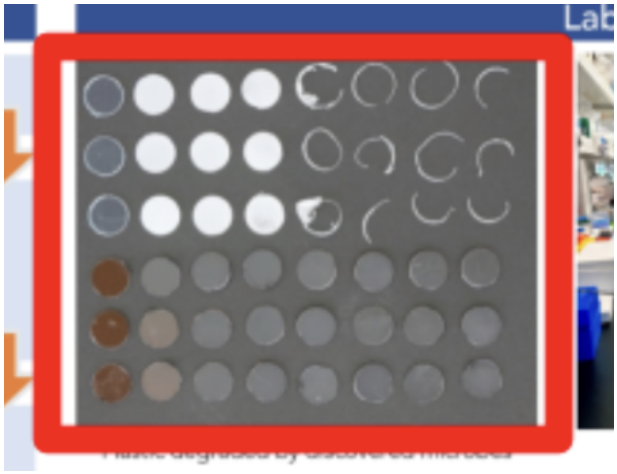
From ProjectBoard online

Update:

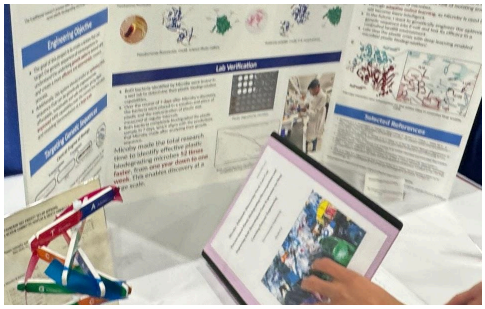


Full, original image has been found in [Biodegradation of different PET variants from food containers by Ideonella sakaiensis | Archives of Microbiology].

This figure from Walter et al. (2022), above, exactly matches the student's figure, below:



Student's figure shown on board + presentation crops out left side showing the actual plastics used (not PHA or PHB like he claims).



Mirrored/faked plastic degradation figure shown as a prominent result on the in-person physical poster board at Regeneron ISEF 2024

If one compares the shapes of the plastics as seen above, they will quickly notice the Regeneron ISEF Finalist's figure is a mirror image of the past published image. This is deliberate data fabrication and **research fraud**.

The image was not only altered by mirroring it, but was combined with another image by the finalist. This goes to show this is not an accident — the Regeneron ISEF finalist intentionally manipulated the image to play it off as their own creation and result. They also placed an image of themself right next to this result, making it seem like it was their own data.

The other graphs are also taken from online and can easily be found via reverse image search. Proof of this is discussed later.

Similar cases of research fraud are discussed in *Science*:
https://www.science.org/content/blog-post/stop-hocusing-your-western-blots-maybe

2. Stealing past researcher's device and claiming as own

The researcher claims to have created a near infrared spectrometer in the project:

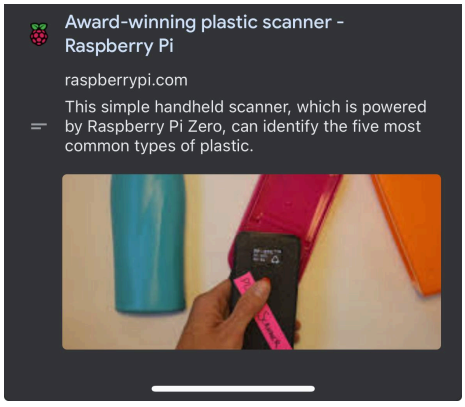
- Using a computer vision-based approach, I created a low-cost handheld near-infrared spectrometer that can analyze the spectral content of a plastic and determine its composition.

Excerpt from the researcher's slides online

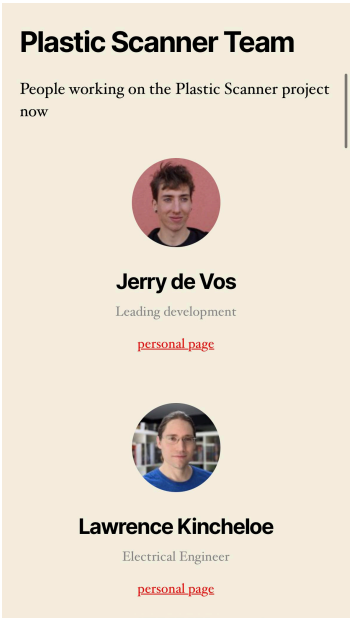
Images used in the finalist's slides and research paper

Yet, they did not "develop" or "create" or "buil[d] a custom sensor." This sensor was not made by the student originally. Using **Google Image reverse search**, one can quickly find that the device the researcher claims to have made was made by Jerry de Vos in 2021:

Article from 2021:
https://www.raspberrypi.com/news/award-winning-plastic-scanner/

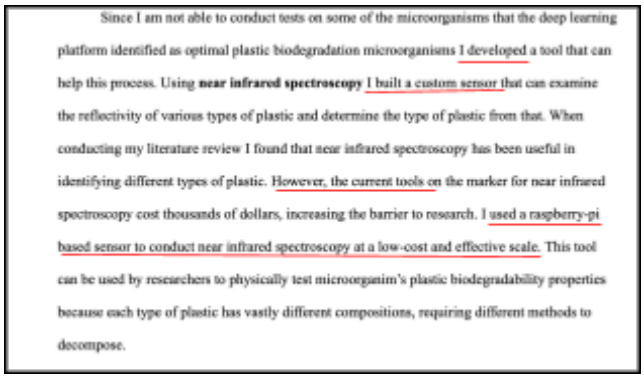


Original image and article published in 2021



Please visit the above webpage for more the original source of the images that the finalist stole.

The researcher of the ISEF project **does not give any credit** to the original creator of the device, or its images and acts like they "built a custom sensor" and "developed [the] tool." **This is very clearly an attempt to conceal the original creator, and for the Regeneron ISEF finalist to act like they developed this solution.**

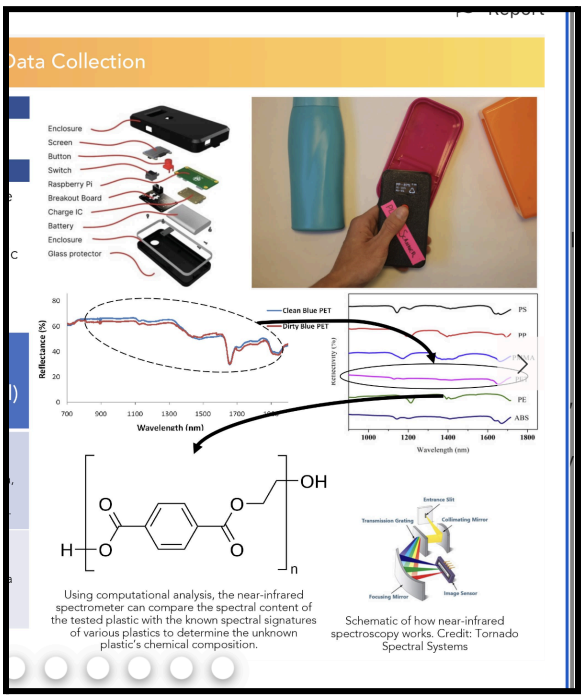


Since I am not able to conduct tests on some of the microorganisms that the deep learning platform identified as optimal plastic biodegradation microorganisms I developed a tool that can help this process. Using **near infrared spectroscopy** I built a custom sensor that can examine the reflectivity of various types of plastic and determine the type of plastic from that. When conducting my literature review I found that near infrared spectroscopy has been useful in identifying different types of plastic. However, the current tools on the market for near infrared spectroscopy cost thousands of dollars, increasing the barrier to research. I used a raspberry-pi based sensor to conduct near infrared spectroscopy at a low-cost and effective scale. This tool can be used by researchers to physically test microorganism's plastic biodegradability properties because each type of plastic has vastly different compositions, requiring different methods to decompose.

Excerpt from the researcher's research paper online - very clearly makes it seem like they developed the device, when it was in fact developed by a different researcher 3 years ago.

Further, as stated in their presentation:

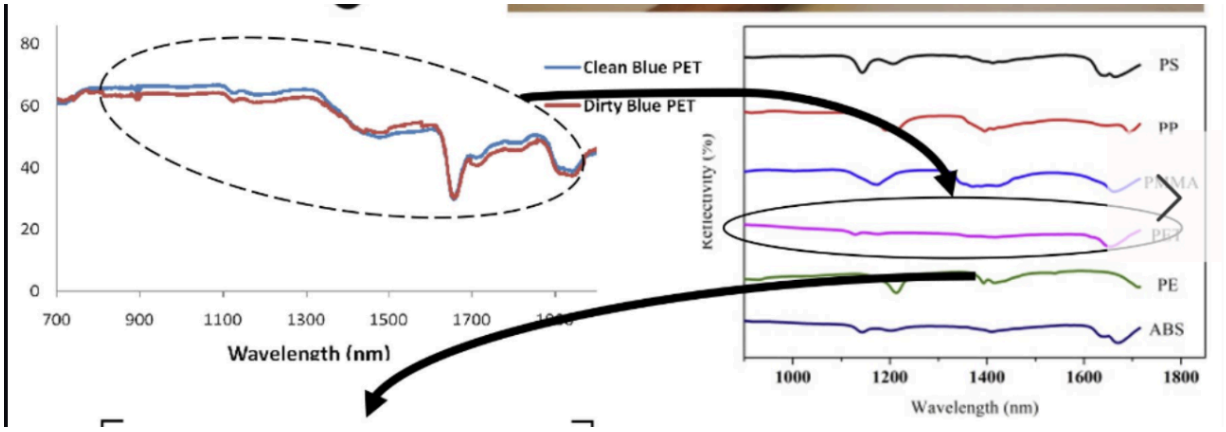
Unless otherwise referenced all images, charts and figures were created by finalist



There is no citation on the poster/slides. The finalist claims to have built the device themselves, but is using images from the original creator without clarifying that they didn't actually invent/develop this device. This is clearly falsification of results.

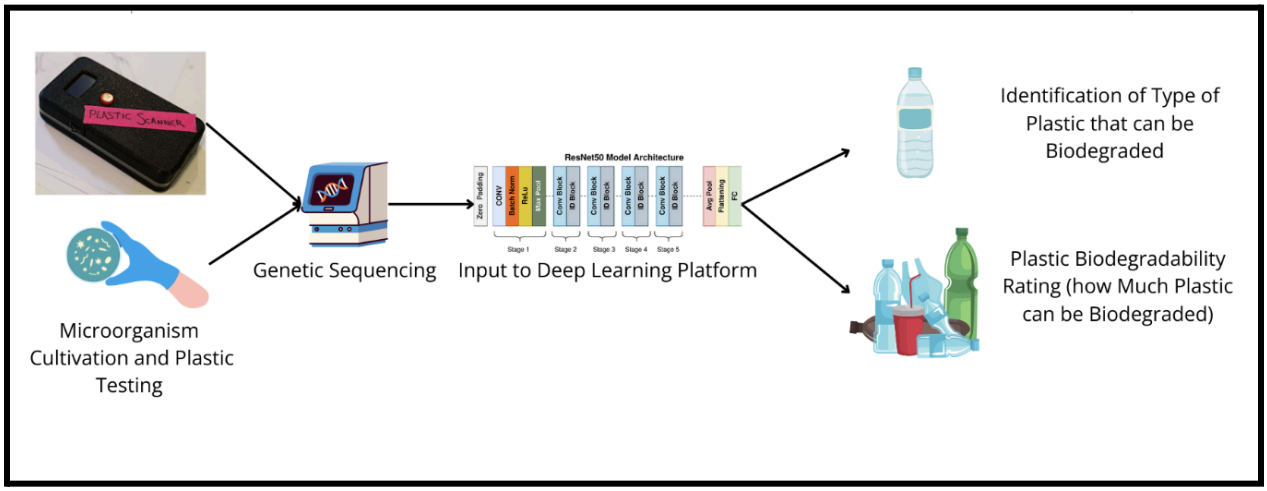
Image once again used on researcher's board, they claim they made this device when they in fact did not.

Further the infrared spectrometer device being shown is actually a multispectral detector. It shines LEDs at different wavelengths and measures the reflection. Thus, the measurement should only have a few discrete points, but the measurements being shown are continuous spectra:



Note: The researcher does credit other images, but conveniently leaves out the credit for the device that they claim to have made, when the images are clearly taken from the person who originally invented the device. Thus, these are not "accidental" citation issues. This is a bigger issue as it is one that the D&S inspectors at ISEF cannot solve — if a finalist is intentionally not labeling certain figures so that people think it is their own.

However, in this case, it is clearer that the finalist did not actually build the device, and faked this part completely. This explains why their data/results seem fake later on as well — the initial data collection never happened, leading to a chain of falsified results throughout the project.



From researcher's paper, plastic scanner can be seen in top left of diagram

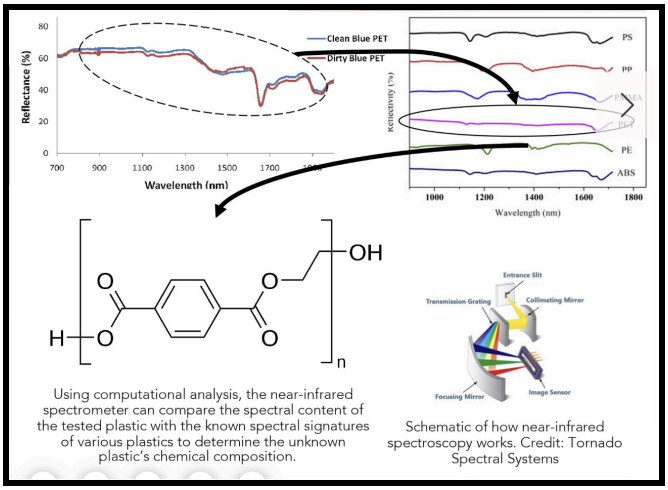
Similarly, in the student's research paper, the plastic scanner is once again depicted here.

Yet, this image is stolen not from the Raspberry Pi article but another article about the original creator: "Nederlandse student wint James Dyson Award 2021 met Plastic Scanner" (link to original source); the student Jerry de Vos from the Netherlands originally had this image put up in this article 3 years ago.

This goes to show the intentional effort the student went through to find images by the original creator across several different articles — and cropped them to make it seem like the student was the inventor.

Side-note: the image of the Deep Learning Platform in the center of the figure depicts a model that has no relevance to the ISEF Finalist's project. The model is ResNet50, which is a CNN architecture utilized in Image Classification, and is not explained anywhere else in the project. Other inaccuracies and other incorrect claims are listed near the end of this report.

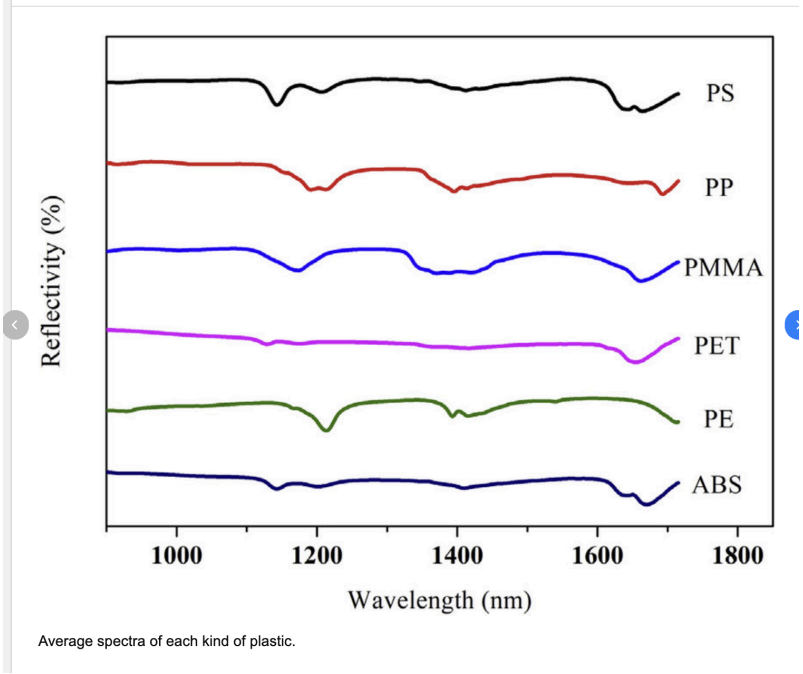
The researcher also steals the wavelength vs reflectivity plot as shown (top right), making it seem like it came from their data:



From the ISEF finalist's ProjectBoard, the figure in question is in the top right.

This figure is right next to a picture of themself to make it seem like it was their own collection and creation.

Original source by Zhu et al., 2019:
[(PDF) Plastic Solid Waste identification system based on Near Infrared Spectroscopy in combination with support vector machine]



Original image from Zhu et al., 2019

The researcher also does not cite this picture, claiming it as their own. They added a few annotations which make it seem like it was their own work, when it was not:

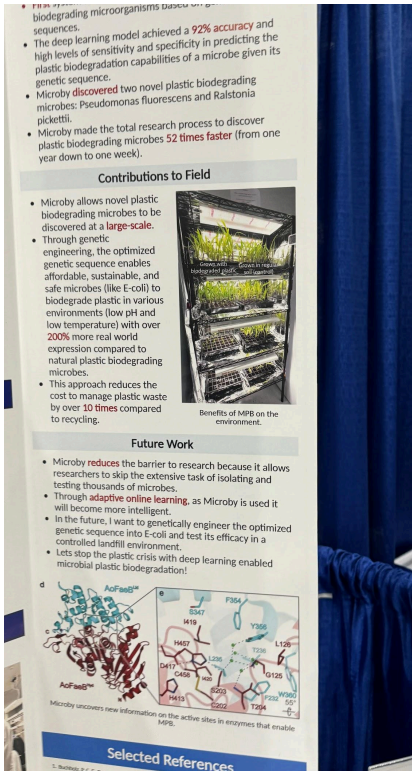


Image of plants grown with white text annotations written on top

These plants were not grown by the finalist. A reverse image search finds this photo online.

3. Claims about novelty

- These two bacteria were not previously found to biodegrade plastic and Microby made a novel discovery with minimal costs and research time.

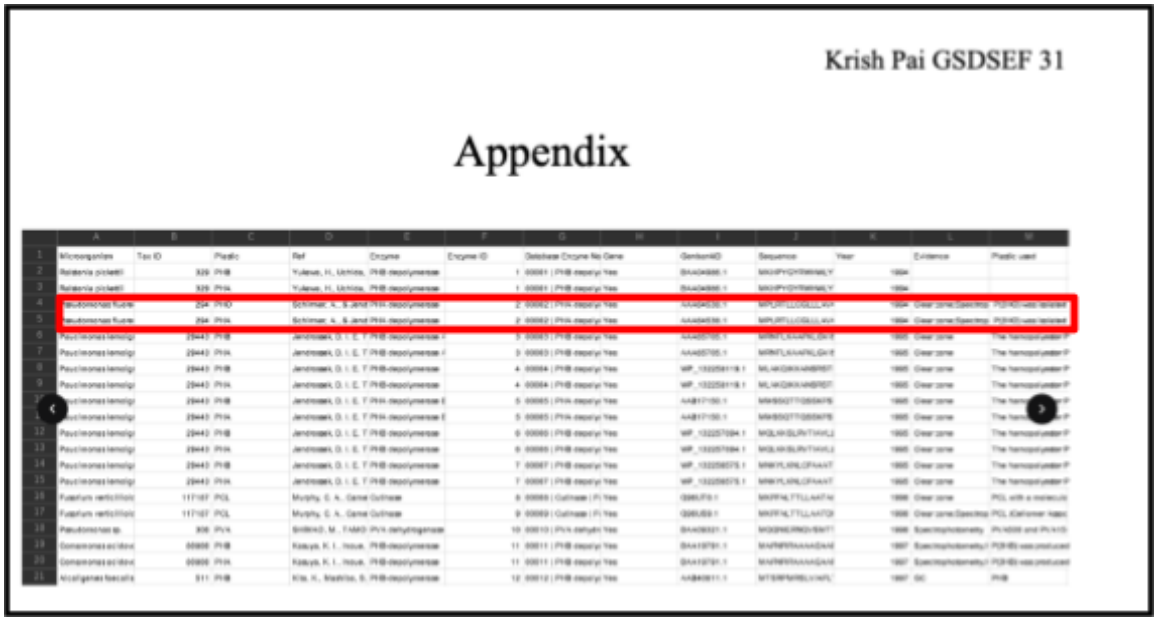
The researcher claims that their deep learning algorithm predicted two bacteria that were not previously found to biodegrade plastic. This is a **central claim of the project**, perhaps the biggest one, and so disproving this would certainly not warrant the student's award.

However, in the finalist's local fair (Greater San Diego Science and Engineering fair (GSDCEF)), they have a diagram in the appendix section that they later do not include in the ISEF presentation. This diagram clearly indicates that there was **past evidence** about both of the

two bacteria, and that the finalist **was aware that these bacteria were previously found to biodegrade plastic.**

The two bacteria the researcher claims to have discovered to have biodegradable properties are *P. fluorescens* and *R. pickettii*.

Yet, in the researcher's **own** Appendix section from their **state level research paper** (they do not have this figure in their ISEF work, probably intentionally), P. fluorescens as shown in red were identified in **1994** to have plastic degrading properties.



<u>The appendix table clearly shows that in 1994, PHA and PHB were identified as degradable by *R. pickettii* and *P. fluorescens* by Schimaer et al. and Ykawo et al.</u>

Yet, the ISEF finalist maintains that they novelly discovered this:

Pseudomonas fluorescens and Ralstonia pickettii bacteria were discovered by the deep learning system to biodegrade polyhydroxyalkanoate (PHA) and polyhydroxybutyrate (PHB) plastics respectively, very effectively.

Microby discovered Pseudomonas fluorescens and Ralstonia pickett as novel plastic biodegrading microorganisms.

These two bacteria were not previously found to biodegrade plastic and Microby made a novel discovery with minimal costs and research time.

In the student's presentation and research paper, they state they obtained their data from PlasticDB and PaZY.

Upon looking at PlasticDB, both organisms are already found in the 20th century to degrade plastic, and **this is the very dataset that the student used** meaning that the discovery of the “novel plastic biodegrading microorganisms” is false:

R. pickettii for PHB plastic:

Tax ID	Microorganism	Plastic	Year	Protein	Thermophilic conditions	Isolation environment	Isolation location	Evidence	Reference
329	Ralstonia pickettii	PHB	1994.0	PHB depolymerase		Soil	Japan		Yukawa, H., Uchida, Y., Kohama, K., & Kurusu, Y. (1994). Monitoring of polymer biodegradabilities in the environment by a DNA probe method. In Studies in Polymer Science (Vol. 12, pp. 65-76). Elsevier. (Google Scholar)
329	Ralstonia pickettii	PHB	1999.0	PHB depolymerase	No			Weight loss	Kasuya, K. I., Ohura, T., Masuda, K., & Doi, Y. (1999). Substrate and binding specificities of bacterial polyhydroxybutyrate depolymerases. International journal of biological macromolecules, 24(4), 329-336. (Google Scholar)
329	Ralstonia pickettii	PHB	2017.0	No	No			Clear zone	Suzuki, M., Tachibana, Y., Kazahaya, J. I., Takizawa, R., Muroi, F., & Kasuya, K. I. (2017). Difference in environmental degradability between poly (ethylene succinate) and poly (3-hydroxybutyrate). Journal of Polymer Research, 24(12), 217. (Google Scholar)

P. fluorescens for PHA plastic:

Tax ID	Microorganism	Plastic	Year	Protein	Thermophilic conditions	Isolation environment	Isolation location	Evidence	Reference
294	Pseudomonas fluorescens	PHA	1994.0	PHA depolymerase	No			Clear zone, Spectrophotometry	Schirmer, A., & Jendrossek, D. (1994). Molecular characterization of the extracellular poly (3-hydroxyoctanoic acid) [P (3HO)] depolymerase gene of Pseudomonas fluorescens GK13 and of its gene product. Journal of bacteriology, 176(22), 7065-7073. (Google Scholar)

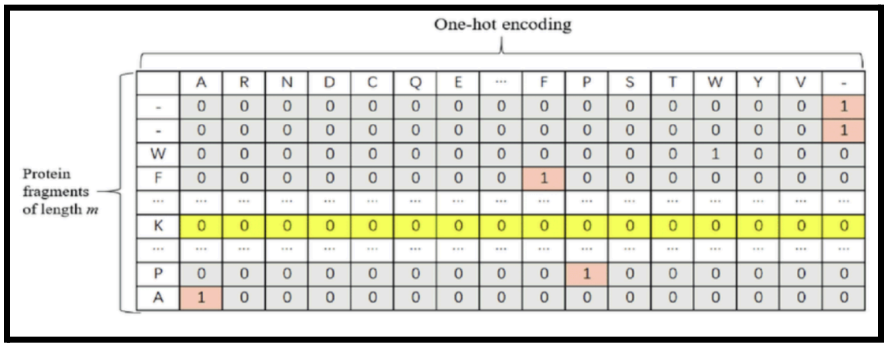
This is the same data shown in the Appendix of the finalist's own research paper. This means that the finalist was already aware that these two organisms had past proof biodegrading PHA and PHB, yet they claim their model "predicted" this fact.

Potential Issues: Plagiarism and Citation Fraud

This section requires further investigation by Society for Science, Regeneron, ISEF, and the community.

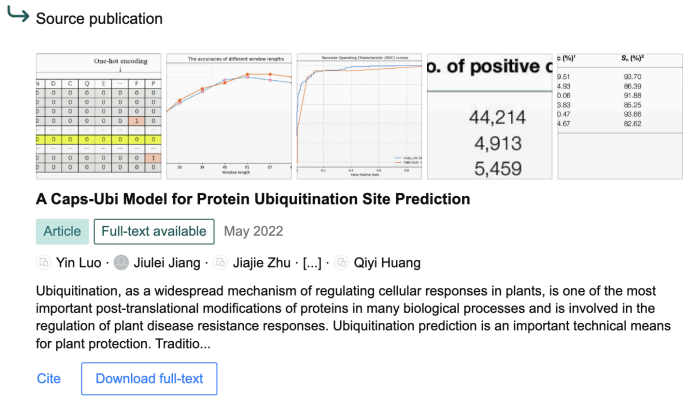
Stolen figures

The ISEF finalist uses the following figure in their research paper.



But this was stolen/screenshotted from online and is not by the student.

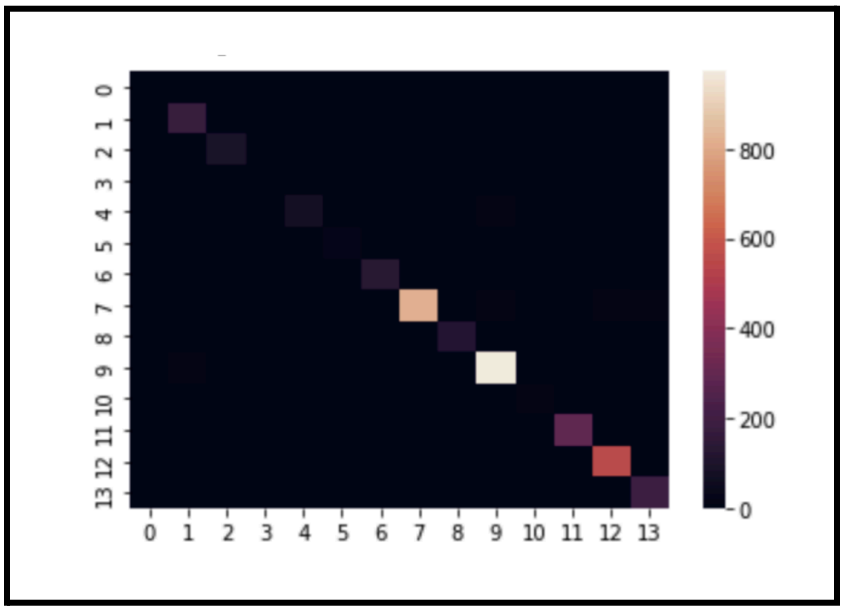
This above figure, found via Google Reverse Image Search, was originally published by Luo et al., in 2022: https://www.researchgate.net/publication/360541857_A_Caps-Ubi_Model_for_Protein_Ubiquitination_Site_Prediction/figures?lo=1



Hence, the use of the word 'protein fragments' -- which does not even have relevance in the Regeneron ISEF finalist's project in which *genes* were used, NOT protein fragments.

The figure is not credited at all, and is a core part of the finalist's explanation of methods.

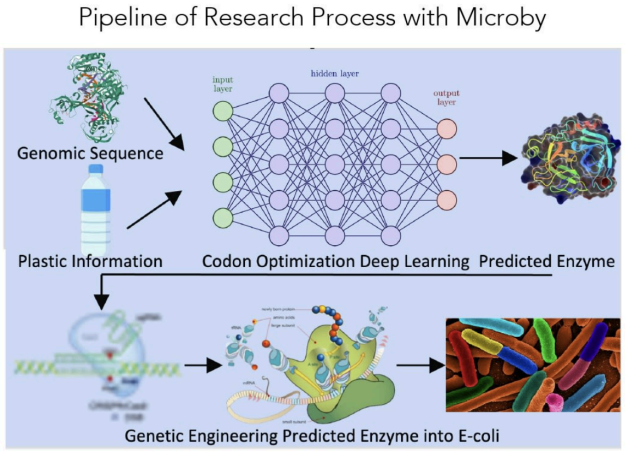
Additionally, the following figure can also be found in the finalist's research report:



1) There is no explanation of what this figure represents
2) This figure is blurry compared to other figures the finalist has made. This suggests that a screenshot was taken
3) There is a small black line at the top of the figure, also suggesting a screenshot was taken, and some wording was cropped out at the top. However, the bottom of one of the letters was left, leaving a noticeable, tiny black line.

Further investigation needs to be done on this figure to determine where the screenshot was taken from.

The other plastic infrared spectrometer figures were also taken, as aforementioned in the earlier section of this report.



On slide 10 of the researcher's slideshow, they have these 7 images. All 7 are taken from online, and none are credited. They make reference to their use of Google Alphafold, which makes it seem like they modeled proteins by themself. However, the images they used were lifted from online sources, and this is easily verifiable using Google Images (reverse image search tool).

This is not a big issue. Nearly every ISEF Finalist has to go through the process of crediting google images, etc. However, the bigger issue is taking actual figures and not crediting them, as aforementioned.

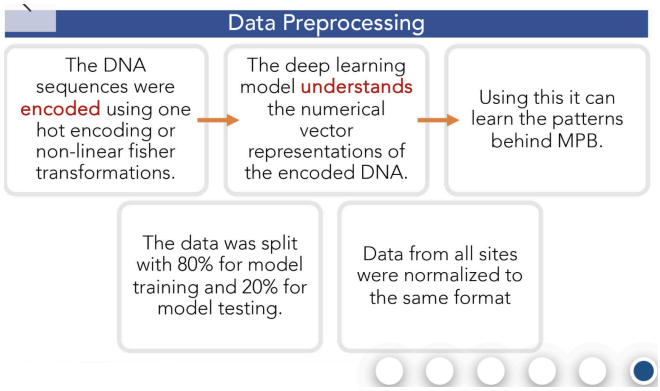
Description of image	Original/Actual source
Protein in top left of researcher's figure	https://www.rcsb.org/structure/4n41
Water bottle image	https://www.beveragedaily.com/Article/2020/07/17/Nestle-Waters-launches-100-rPET-bottles-for-three-US-brands
Neural network image	https://www.researchgate.net/figure/Model-approximator-architecture-the-hidden-layers-approximate-the-dynamics-transition_fig1_379422460
Trypsin protein image	https://en.wikipedia.org/wiki/Trypsin
CrisprCas9	https://www.mdpi.com/horticulturae/hortic

	ulturae-07-00193/article_deploy/html/images/horticulturae-07-00193-g001.png
Ribosome	https://commons.wikimedia.org/wiki/File:Ribosome_mRNA_translation_ku.svg
Colorful rod shaped bacteria	https://news.harvard.edu/gazette/story/2014/12/bacteria-churn-out-valuable-chemicals/

Once again, this issue is not necessarily that the author didn't credit some of these images. This happens to nearly every ISEF finalist by mistake. This is possibly a genuine mistake that was overlooked by Regeneron ISEF D&S. **It's that they screenshotted actual figures from real research papers and past authors and played it off as their own**.

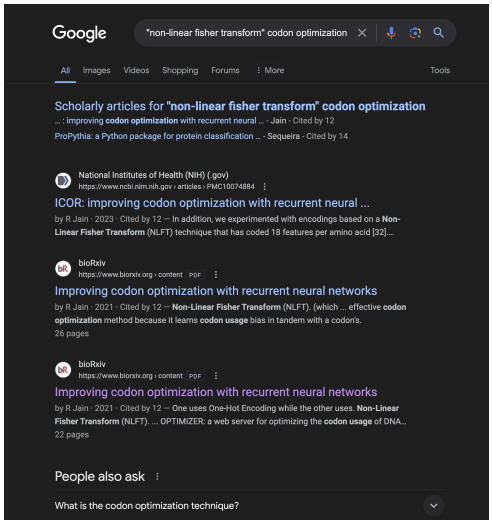
Similarities to past winning ISEF project

According to https://projectboard.world/isef/project/enev081-efficiently-discovering-plastic-degrading-microbes, this project's ProjectBoard website, in the 'project presentation' tab, the following claim is made on the presentation for this finalist's project:



The piece to highlight here is in the first box, stating "DNA sequences were encoded using one hot encoding or non-linear fisher transformations."

If one searches up "non-linear fisher transformation" related to codon optimization, the only result that comes up is from a [past ISEF winner (2022) whose project was on codon optimization]. There are **many, many similarities to this past winner's project** throughout the new winner's project, suggesting that the new winner took major inspiration, but worse, did not credit the things that were taken from the past winner's project and subsequent research paper.



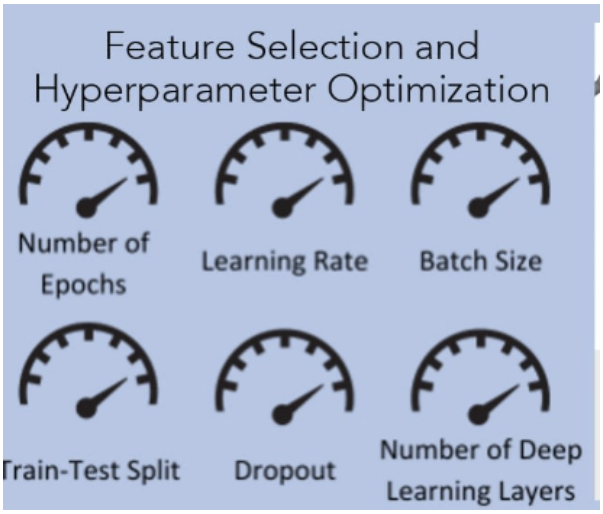
The non-linear fisher transform is simply a mathematical technique. The specific implementation of using it to encode amino acids means you must need some features of the amino acids themselves. This was developed by Jain et al. in their 2023 paper.

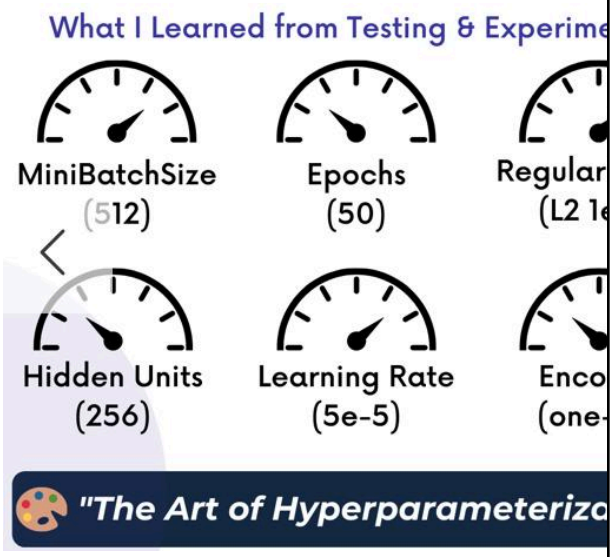
Yet, this year's Regeneron ISEF winner **does not cite this past paper** or make any reference, making it seem like either:

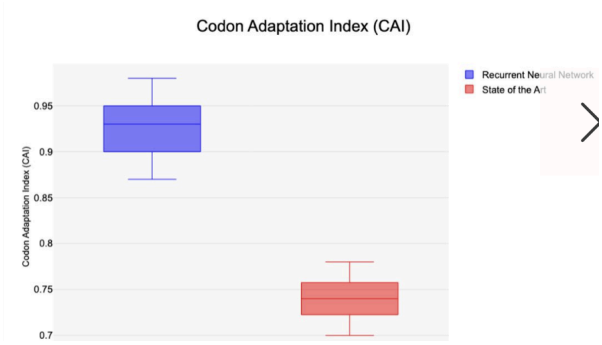
1. The finalist did not actually implement the non-linear fisher transform technique
2. OR, they lifted code/algorithms from a past research paper without attribution.

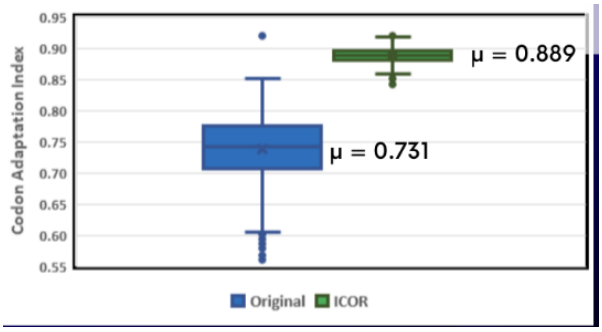
Codon optimization is a very niche problem, so it is unlikely for two researchers to go down this realm at ISEF, and for both to win top awards. Inspiration is understandable, however, there are some elements that are directly lifted from the past winner's project, as also mentioned in a later section of this report:

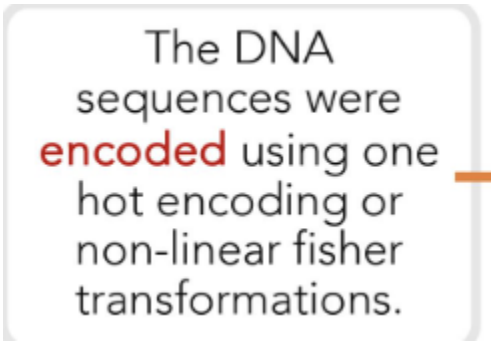
Seemingly copied/inspired material from Regeneron Young Scientist Award winner (2024)	Material from winner of Regeneron Young Scientist Award (2022)
Uses a recurrent neural network for codon optimization on "over 6 million genes" achieving a "**over 200%** increase"	Uses a recurrent neural network for codon optimization (specifically shows biLSTM architecture) on "6.8 million genes", of which "40,000 non redundant" and "7,000 high expression" are used achieving "**236%** increase"



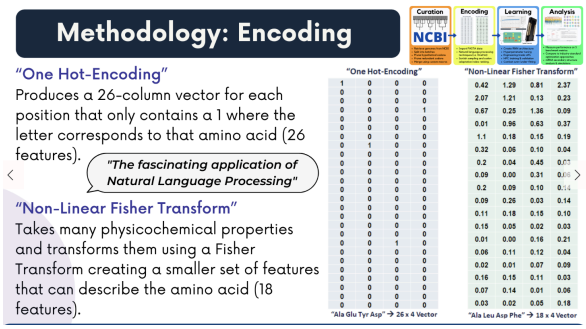






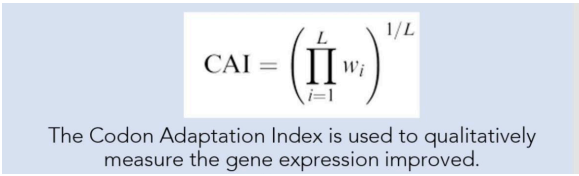


The 2024 ISEF finalist does not show their data of the physicochemical properties that are required to perform a fisher transform. Thus, it appears that they simply copied this terminology from the 2022 Finalist without actually performing the transformation.



This ISEF finalist's published research paper is the only to make reference to non-linear fisher transform in the context of codon optimization.

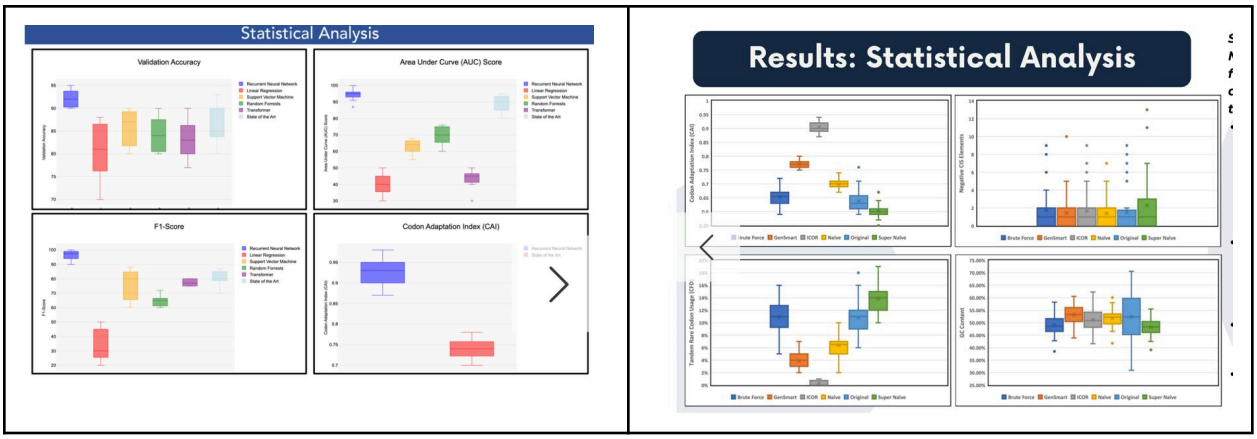
<table>
<tr>
<td>over 6 million genes.</td>
<td>Wrangled 6.8M+ genes to assemble robust data for deep learning.</td>
</tr>
<tr>
<td>improved CAI by 0.2, indicating over 200% more real-world gene expression. This enables genetically</td>
<td>Vaccine Manufacturing:
Produce over 2x more vaccines for same cost & time. Apply to malaria,

Based on improvements to CAI, ICOR yields a ~236% and ~28% improvement in protein expression</td>
</tr>
<tr>
<td> 

Here, the 2024 ISEF winner took a screenshot of the 2022 ISEF winner's formula for CAI. It is not a figure created by the 2024 finalist because it is blurry and has a white background.

Further, upon reverse image search, there is no source online except the 2022 ISEF winner's project that has the exact same formatting. The 1/L exponent is a clear marker of this as all other CAI formulas online have this as an actual fraction not a forward slash. There was little reason to screenshot someone else's project formula, but it clearly shows the 2024 Winner was directly looking at the 2022 Winner's project — which is more evidence in the direction of the 200% and 6 million gene number also being falsely copied.

Also, they mistakenly said "qualitatively" — likely a typo after copying the 2022 project's wording.</td>
<td>$$CAI = \left(\prod_{i=1}^{L} w_i\right)^{1/L}$$
$= 0.889 \pm 0.012$
Codon Adaptation Index is used to quantitively measure gene expression because the reference set is made of highly expressed genes.</td>
</tr>
</table>



This table is just to serve as evidence that the finalist was clearly aware of the past paper/project (due to multiple visual elements being similar), yet the work was not cited + there are extremely similar numbers being claimed in terms of genes and performance which is odd, while the new project gives no reference to literature that allows one to quantify such performance.

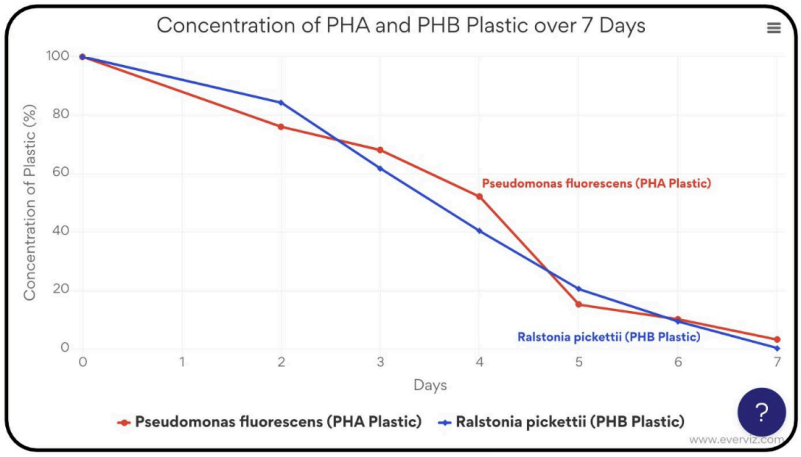
This is considered citation fraud when you borrow a novel methodology (i.e. NLFT for codon optimization was a novel method created by the 2022 Finalist) but doesn't cite the method anywhere or the use of it. Also, even if they were just inspired by the project, proper research practice would be to cite the published paper (https://doi.org/10.1186/s12859-023-05246-8)

Any one of these observations in the table above alone could be coincidental, but because there are so many similarities, it is possible that research plagiarism may have occurred in which the 2024 Finalist stole methods and innovations from the 2022 Finalist's research, without giving any due credit.

Scientific/Factual **potential** inaccuracies:

This must be prefaced by saying these are unverified parts of the project that could point to data falsification, but require further investigation on part of Society for Science and Regeneron.

In the project, the main claim is that 100% of plastic is degraded.



Both microorganisms identified by the deep learning system completely biodegraded plastic in 7 days.

100% degradation claim is made (images of plastic degradation were already shown to be faked/stolen in the earlier section)

The researcher claims that the same neural network model that performs codon optimization ALSO predicts microbes that will degrade plastics effectively.

They claim that they predicted **two novel microorganisms**. This is completely false, as the researcher's own paper from their local fair disproves (see above section about novelty).

Yet, there is another issue. They claim that within 7 days, **both bacteria** degrade 100% of plastic. According to past literature, this number is unfeasible:

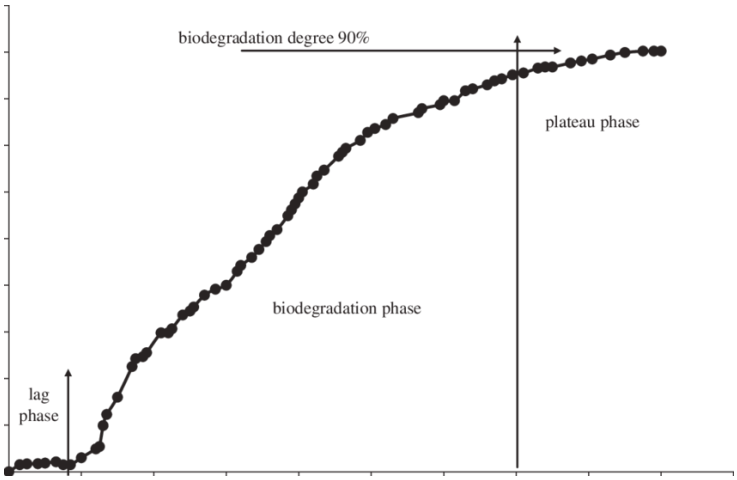
- https://pubmed.ncbi.nlm.nih.gov/15162769/#:~:text=Efficacy%20of%20the%20microbial%20species,plastics%20in%20one%2Dmonth%20period.

This paper tests Pseudomonas species and finds "Efficacy of the microbial species in degradation of plastics and polythene was analyzed in shaker cultures. Among the bacteria, Pseudomonas species degraded 20.54% of polythene and 8.16% of plastics in one-month period."

https://www.sciencedirect.com/science/article/pii/S2667010021000354#:~:text=It%20is%20indicated%20that%20many,at%2030%E2%80%9337%20%CB%9AC.

- This paper tests the exact same species that the ISEF finalist used, and found **6.25% degradation after 30 days**.

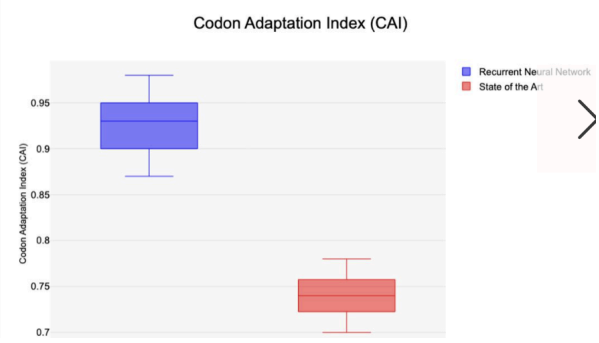
Thus, claiming that somehow these bacteria will degrade 100% of plastics is nearly impossible.



As it can be seen in past literature, there is a lag phase and a biodegradation degree, at which all plastics would cease to be degraded. **This looks like a logistic type of growth.**

Yet, the finalist somehow obtains a **linear graph** in which the plastics are degraded 100%.

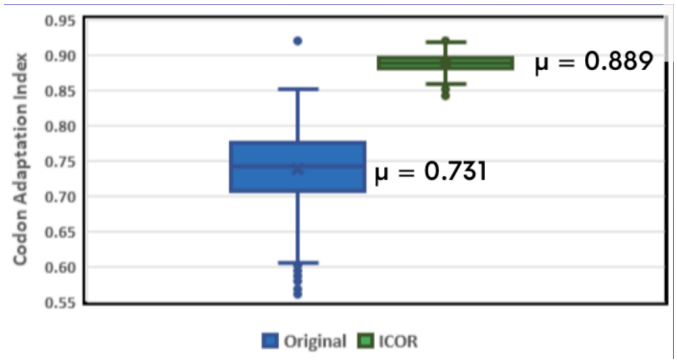
Furthermore, this also debunks the novelty claim once again, as the bacteria have already been lab tested in past papers. Thus, it is not fair the finalist to say: “discovered two novel.”



The researcher's main chart about codon optimization improvements is this one, showing ~20% "improvement" in CAI. Nowhere in the project material do they explain what the "state of the art" is.

The error bars for the "state of the art" are also likely falsified. As the researcher points out in their video and slides, etc. -- they are comparing their deep learning approach to the "original microbe"'s genome. What is the CAI of a genome of a usual microbe?

According to past literature, the CAI of microbial genomes is much more variable (see error bars or whiskers on the blue box below):



Jain et al., 2023 (also the 2022 ISEF Regeneron Young Scientist Award Winner who did a project on codon optimization): [*Link to source*]

This is not simply an experimental error or difference on part of the researcher. They are looking at **known** organisms with **known genomes**. Thus, the CAI calculated should not be variable in a new measurement.

E.coli	
gene	CAI
17 RPs	0.467-0.813

Sharp & Li, 1987

Just 17 E. coli genes had CAI ranging from 0.467 to 0.813!

Thus, the ISEF researcher using 6 million genes and having such a small error bar/whiskers points to falsification / incorrect error bars.

Even if the 2024 ISEF Finalist used a different subset than what past researchers used, there should still be a larger variation in CAI for reference genes, which would result in larger/longer error bars and whiskers, as well as outliers.

There are no outliers in the 2024 Finalist's data on CAI, and all the results are shown to be normally distributed, which is oddly suspicious.

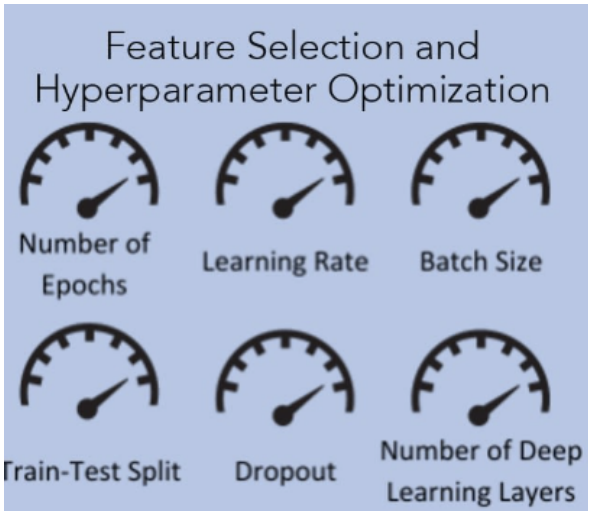


Image Depicting Feature Selection and Hyperparameter Optimization

In this image, the finalist claims that train-test split was a step in their feature selection and hyperparameter optimization, but train-test split is not a part of this step in model development. This is a minor factual error.

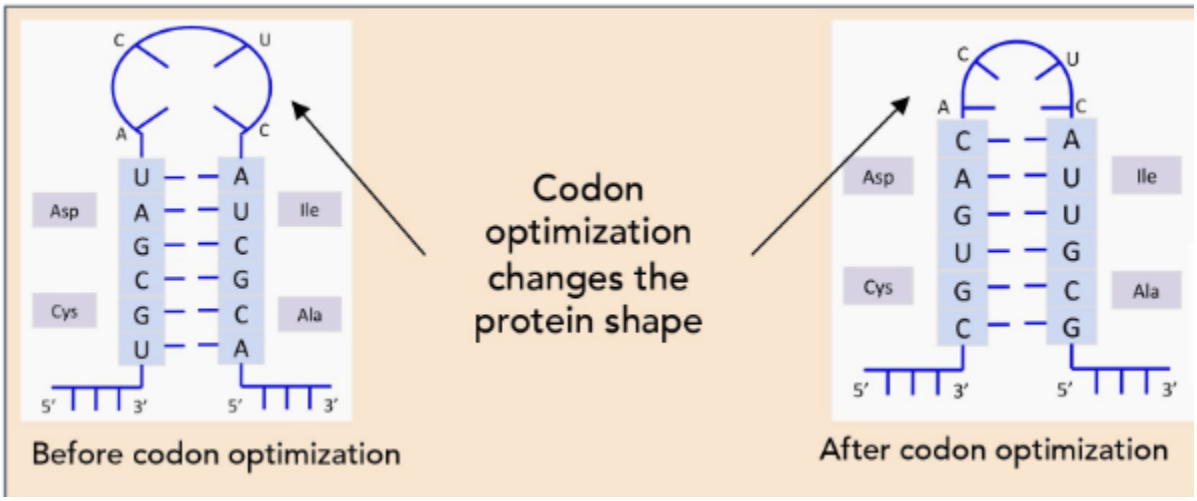
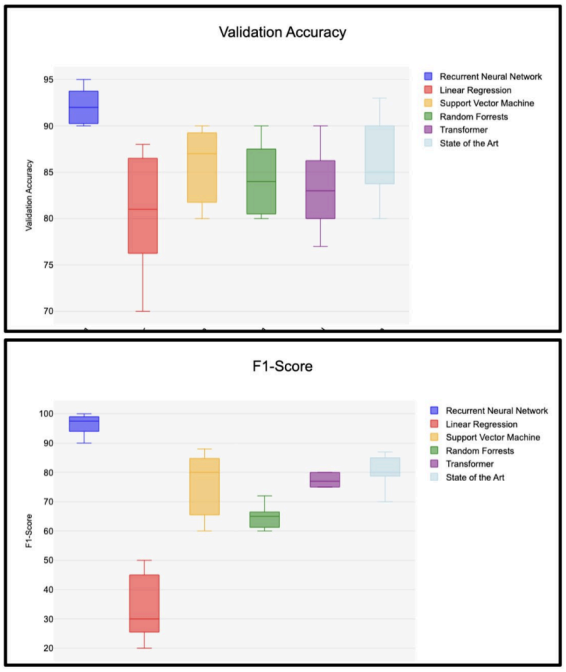


Image by researcher claiming that codon optimization changes protein shape

The depicted images are mRNA, not proteins, as the stem-loop is shown. Further, codon optimization would alter this hairpin shape, which is not linked to the end protein conformation. Past literature finds shape is changed because of other reasons (i.e. rate of elongation is different, thus causing conformational changes). Thus, the use of these images with the claim that codon optimization changes protein shape is incorrect. This is a minor factual error.



Having a box plot for F1 score and validation accuracy fundamentally misunderstands what these metrics represent.

Validation accuracy is a metric that tests accuracy of a machine learning model during training. Validation would be one value over the batch tested — an average of let's say 90%. **There is no variation** where a box plot is necessary, there should just be one datapoint.

F1 score takes error rates across all of the testing data points. Thus, there is **no variation in the data** and there would not be a box. This important distinction, combined with suspicious elements below, along with the earlier point about error bars being falsified, contribute to a strong chance that this data is not real.

The reason this student used box plots is because the student copied the style from the 2022 ISEF project that also did codon optimization; yet, that student compared different metrics which do have variation. F1 Score and Validation accuracy do not.

This is a major technical inaccuracy that points towards falsification of results.

Further, the machine learning models that are claimed to be trained do not make sense either.

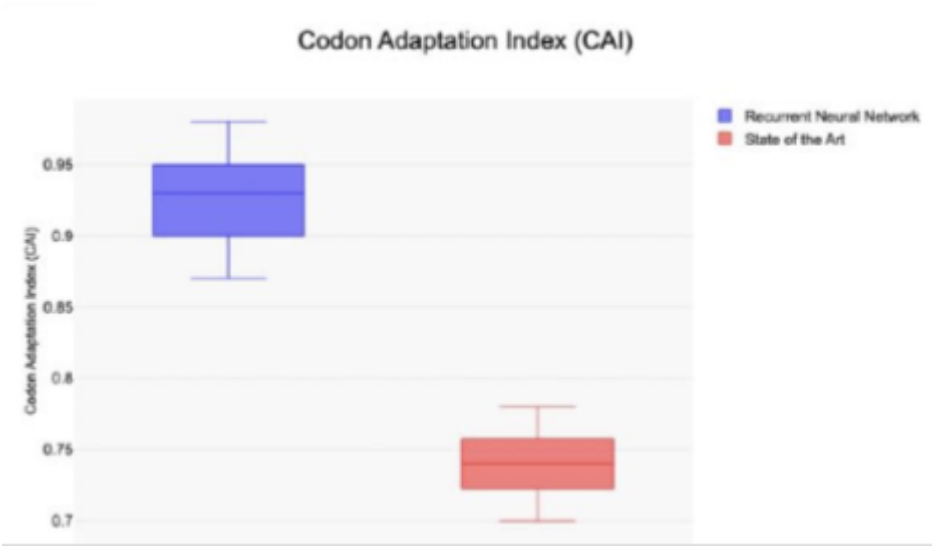
For one, how can a linear regression algorithm be trained on sequential amino acid sequences? This cannot fit into a y=mx+b form. This would require generalized linear models (GLMs), not linear regression. Thus, it does not seem possible to use data with the type that the finalist did towards a linear prediction, when the data is, in fact, very much non-linear.

Similar concerns can be raised about the random forests algorithm being used for this problem.

Support vector machines are known to have very poor performance on large datasets. With a claimed dataset of 6M genes, it would make sense for the transformer to perform significantly better than all the ML methods. Transformers are the state-of-the-art method for sequential data, as evidenced in all literature. Yet, this is not observed in the finalist's project. This is a very suspicious result, and machine learning experts tasked by Regeneron ISEF need to thoroughly review this.

The student likely did not realize what they were claiming is impossible as they also state RF, Linear Regression, etc. to be "deep learning methods" which they are not. These are standard machine learning techniques; deep learning techniques use neural networks.

Furthermore, the "recurrent neural network" approach shows up to 100% F1 Score. If this were the case, there would be at least one instance of 100% validation accuracy as well (as this is a classification task). Yet, the validation accuracy only goes up to 95%. This is another fundamental issue in the data that points towards falsification.



Codon Adaptation Index boxplot

One of the biggest claims made in this project is that the resultant tool of the research, Microby, "improves genetic expression by over 200%." Specifically, the researcher states their tool "improv[es] the **real-world protein expression** by 200%."

Yet, there is no lab testing of real-world protein expression being performed. The 200% figure is based on the Codon Adaptation Index (CAI) chart, as shown above.

Yet, codon adaptation index simply compares the frequency of highly used codons in a reference organism to the sequence provided. There is no reference to which sequences are used as comparison, but, because the researcher claims 200% improvement in a potential E. coli system, they must be referring to E. coli genes.

Further, it is extremely odd that the codon adaptation index (CAI) of the recurrent neural network tool has a mean of 0.93 approximately. The tool is trained on a dataset of 6 million genes, yet these are not high-CAI genes. Thus, it makes no sense for the resultant tool to raise the CAI to such a high number. If their system is trained on E. coli genes to learn patterns from them, how can it have such a large improvement over the very genes it was trained on? A model cannot generalize outside of training data.

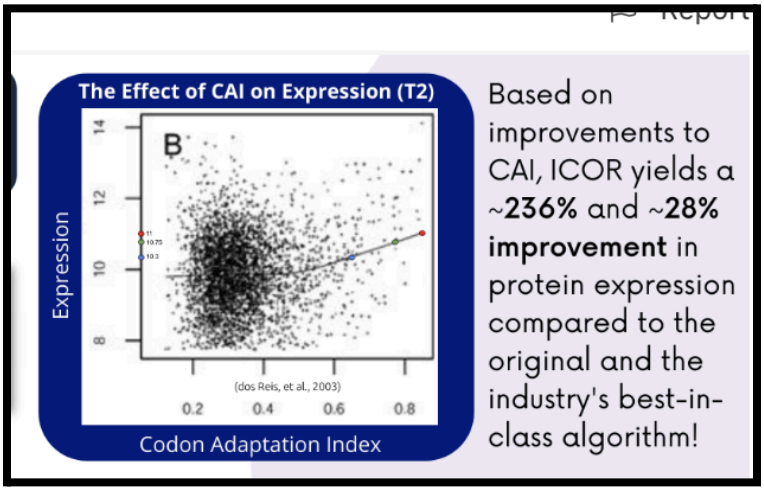
In past research, high-expression genes are trained on. Yet, Microby does not do this. Thus, an increase in CAI points to potential falsification of results, especially to such high levels (0.9+).

However, even this does not make sense. They say that the Microby tool would increase protein expression of the protein degrading enzyme by 200%. But that enzyme does not naturally exist in *E. coli*, so what are they comparing expression/CAI to?

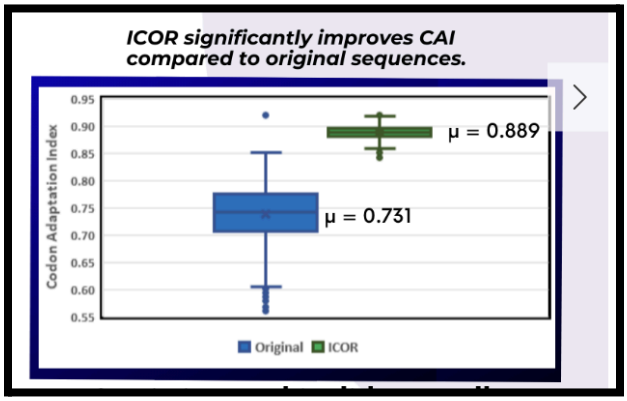
In a separate but related direction, CAI is not equivalent to protein expression. There is no literature that supports an increase of 0.2 in CAI leading to a 200% "real world" expression increase as claimed:

> Measuring the codon adaptation index (CAI), the deep learning system's identified optimal genetic sequence improved CAI by 0.2, indicating over 200% more real-world gene expression. This enables genetically

The researcher is likely basing this number on https://projectboard.world/isef/project/enbm074---synthetic-dna-engineering-with-icor, a past winning ISEF project that has many other similarities to this finalist's work:



Screenshot from past winning project



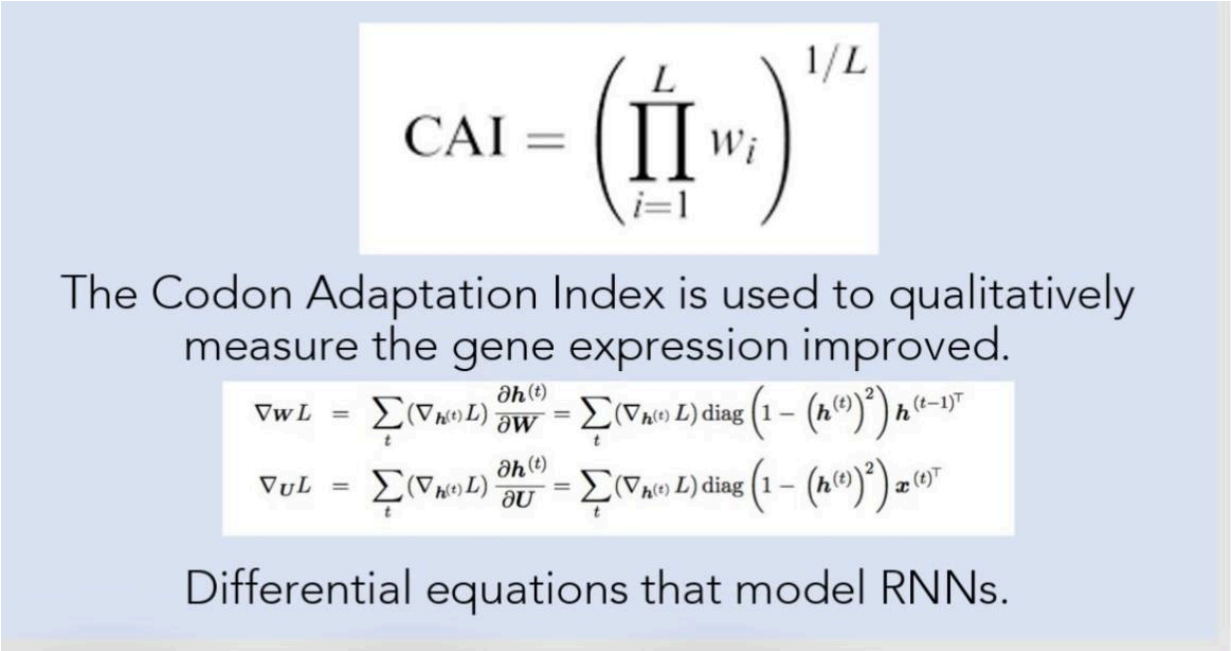
Screenshot from past winning project

The prior winning ISEF project (referred to from here on as “Prior Work”) calculates an exact increase in CAI of 0.158, which corresponds to a hypothetical improvement of 236% based on dos Reis et al.’s expression graph.

Yet, there is no reference to dos Reis et al. or Prior Work in this year’s winning ISEF project. In fact, there are 0 references to codon optimization papers at all. This is extremely suspicious because without sources, how is a number of 200% improvement being calculated over “state of the art”?

So where are the numbers being calculated? The claim of real-world expression makes it seem like there were wet lab results, yet no gel or any results are shown. Instead, strange, ballpark round numbers of “200%” are claimed without any statistics or specific details.

This, along with the strange error bars as posited earlier point towards data fabrication.



CAI is a quantitative measurement, not qualitative. This is an important distinction.

The researcher screenshotted the formulas from https://dev.to/rimmel_codes/recurrent-neural-networks-rnns-easily-explained-5amm and did not cite the original author who derived these formulas! This derivation is extremely long and not very common — this should be cited.

Furthermore, it is a slightly technical misunderstanding for the Regeneron ISEF Finalist to say these differential equations model RNNs. For one, these are partial differential equations. Two,

these are for optimizing parameters of the RNN—not to model it. If it were modeling the RNN, then what is the point of the RNN? You could just PDE-solve the whole problem rather than using neural network black boxes.

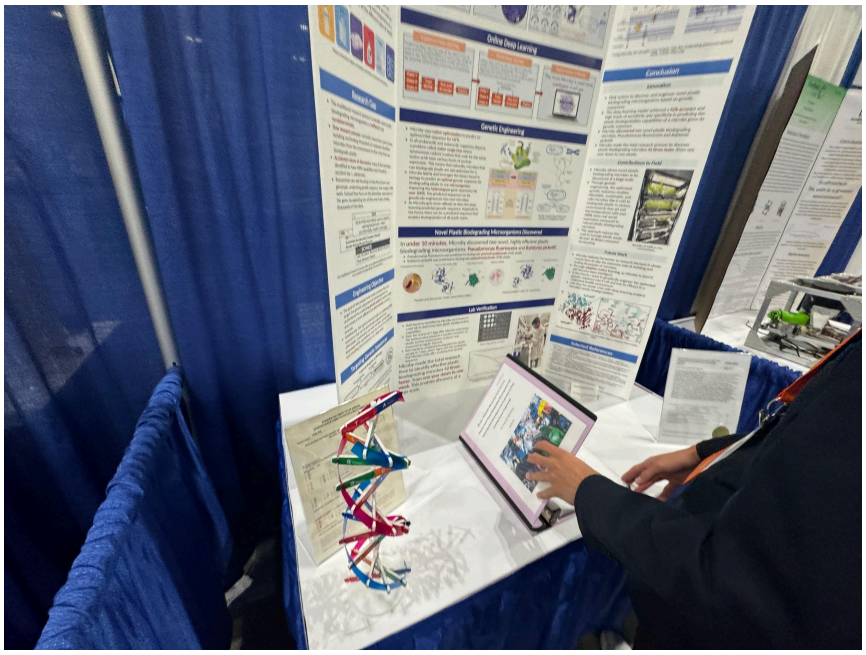
This points to a lack of underlying knowledge on the subject which led to hasty copy pasting from past work. This is also shown in the faking of machine learning box plot data.

This requires future investigation.

The student uses this image to claim wet lab proof of the work. It's important to note:

1) There is nothing related to the project shown in the image
2) Further, there is no tip on the pipette and the vial shown is closed
3) You don't even touch your pipette until you open the solution lid (which is water in this cas) and set the cap down inside contacting the table surface (or sterile hood surface)
4) You don't hook the pipette that way; you hook it by your intermediate phalange not by one's palm.

This is not a major issue, however, it needs to be investigated to find if the student actually performed any experiment in the lab, given that the data was proven to be faked. This requires further investigation, especially because, as shown in the earlier sections, the proof of the microscopy/images of the plastics was taken from online/the internet.



There was no Form 1C shown in physical display board

Conclusions

There is an enormous amount of direct proof, as well as potential issues that require further investigation, that warrant the Society for Science and Regeneron ISEF to take action.

This is to maintain the integrity of the research community and act in the fairness of all. We absolutely, unequivocally discourage the harassment of any members of the ISEF community, including the project discussed in this report. Everyone is human, and humans can make mistakes. However, people need to take responsibility for cheating and unethical practices such as the ones outlined in this report. We should establish a culture of accountability rather than denial → cancellation, which is what is currently going on.

The burden is on Society for Science, Regeneron, and ISEF affiliates to take action. We encourage readers and community members to share this report with those who can take action:

Contact San Diego Fair: https://www.gsdsef.org/about/contact
Emails to send: majmera@societyforscience.org, ISEF@societyforscience.org, society@societyforscience.org, alumni@societyforscience.org, communications@societyforscience.org

Contact Jacobs, sponsor of ENEV category: https://www.jacobs.com/contact/public-relations