TEACHER GUIDE
Chemistry

LEARNING UNDEFEATED
Educate | Elevate | Empower

Alien Elements: Exploring Data Science in Chemistry

Topic(s): Chemistry, Data Science, Computer Science

Grade level(s): 9th – 12th grades

Time: 60-75 minutes

NGSS Alignment: HS-PS1-1, HS-PS1-2

TEKS Alignment: CHEM.5A, CHEM.5C

Virginia Science SOL Alignment: CH.2C

ACTIVITY OVERVIEW

In this activity, students are introduced to the field of data science and how it applies to the studies of chemistry. Students will use the physical and chemical properties of alien elements to group and sort them into a table that conveys information about relevant trends and patterns. Students will also be introduced to ideas of how data science can be predictive through the use of certain analytical tools. In this session, students will model a generative adversarial network (GAN), a type of machine learning that utilizes two neural networks. Students will act as the generators and interact with a discriminator, a computer program that has been trained to identify patterns in a given data set about molecular compounds. Students will use the feedback from the discriminator to propose possible compounds that could exist but are not present in the discriminator's data set. This is an example of synthetic data, a tool that is becoming more and more important in science today.

Table of Contents

ALIGNMENT TO STANDARDS

NGSS:

HS-PS1-1. Use the periodic table as a model to predict the relative properties of elements based on the patterns of electrons in the outermost energy level of atoms.

HS-PS1-2. Construct and revise an explanation for the outcome of a simple chemical reaction based on the outermost electron states of atoms, trends in the periodic table, and knowledge of the patterns of chemical properties.

TEKS:

CHEM.5.A Explain the use of the chemical and physical properties in the historical development of the Periodic Table.

CHEM.5.C Interpret periodic trends, including atomic radius, electronegativity, and ionization energy, using the Periodic Table.

Virginia Science SOL:

CH.2C. Trends within groups and periods including atomic radii, electronegativity, shielding effect, and ionization energy.

LEARNING OUTCOMES

Students will know:

- What data science is.
- Data science can be applied to answer questions in many different fields/industries.
- That GANs are used in deepfakes

Students will understand:

- Trends in the Periodic Table including atomic radius, electronegativity, and ionization energy
- Similar properties are grouped together vertically in the Periodic Table
- How predictive tools, like GAN, can help to produce new data to aid in molecular discovery
- How deepfakes can be harmful

Students will be able to:

- Manipulate and interpret graphs to find trends and patterns in element properties.
- Create a visualization in the form of a table of the patterns and trends identified.
- Model the generator side of a GAN to produce viable new molecular compounds.
- Discuss image, video, and audio GANs and deepfakes and the pros and cons of them.

CAREER CONNECTIONS

Data Scientist

Data scientists use analytical tools and techniques to extract meaningful insights from data.

Work Environment: Data scientists spend much of their time in an office setting. Most work full time.

Duties: Professionals in these jobs have the following duties and more: determine which data are available and useful for the project, create, validate, test and update algorithms and models, use data visualization software to present findings.

Median Salary: $103,500 (US Bureau of Labor, 2022)

Source: https://www.bls.gov/ooh/math/data-scientists.htm

Computer and Information Research Scientist

Computer and information research scientists design innovative uses for new and existing computing technology.

Work Environment: Computer and information research scientists collaborate with engineers or other specialists or research scientists in different locations and do much of their work online.

Duties: Professionals in these jobs have the following duties and more: analyze the results of experiments, design and conduct experiments using techniques from data science and machine learning, explore problems in computing and develop theories and models to address those problems.

Median US Salary: $136,620 (US Bureau of Labor, 2022)

Source: https://www.bls.gov/ooh/computer-and-information-technology/computer-and-information-research-scientists.htm

Database Administrators and Architects

Database administrators and architects create or organize systems to store and secure data.

Work Environment: Many database administrators and architects work in firms that provide computer design services or in industries that have large databases, such educational institutions and insurance companies. Most database administrators and architects work full time.

Duties: Professionals in these jobs have the following duties and more: design and build new databases, make and test modifications to database structure when needed, ensure that organizational data are secure.

Median US Salary: $112,120 (US Bureau of Labor, 2022)

Source: https://www.bls.gov/ooh/computer-and-information-technology/database-administrators.htm

BACKGROUND INFORMATION

In this activity, students will apply data science techniques to understand relationships between chemical elements based on physical and chemical properties. They will do this with alien elements. These alien elements have the same properties as our Earth elements but have been disguised with a new name.

Students will first be introduced to the field of data science and its importance. The goal of data science is to find knowledge and insights from noisy data. Because the goal is so broad, it can be applied to many different fields from science to business. In the activity, students will partake in the exploration, visualization, and machine learning phases of the data science life cycle.

In the first half of the activity, students will use the Common Online Data Analysis Platform (CODAP) to visualize patterns and trends in physical and chemical properties. Students can drag a property to an axis on the provided graph. They can also select certain data points to zoom in on, hiding unselected cases. During this phase of the activity, students should identify which property is the most unique (atomic weight) and should see some properties create overall trends or repeating patterns when compared to atomic weight. With this information, students will then sort and place cards in a table to visualize the patterns they've identified.

In the second half of the activity, students will model a generative adversarial network (GAN). This tool is composed of two neural networks, a generator and a discriminator. The discriminator has been trained to recognize data, in this case, the compounds found on the alien planet, Dod. Students will work as the generator to produce new compounds that could exist. The discriminator is looking for the same patterns as its dataset, so students will need to learn from their responses to find possible answers. Students will be able to propose compounds that include two different elements. Students will also be able to designate the number of atoms present in the molecule with the subscript. When this proposed idea is given to the discriminator, the discriminator will supply 3 pieces of information: (1) the chemical formula for the proposed compound, (2) the valence electron count of the proposed compound, and (3) whether or not the compound passed or failed the discriminator. Students will need to find patterns in the responses from their generated compounds to identify valid compounds.

<u>Resources About Data Science</u>
What is data science?
- https://www.youtube.com/watch?v=RBSUwFGa6Fk

What is data science?
- https://ischoolonline.berkeley.edu/data-science/what-is-data-science/

25 Data Science Applications and Examples
- https://builtin.com/data-science/data-science-applications-examples

Data Scientist Helps Inform and Imagine Army's Workforce of the Future
- https://www.army.mil/article/266979/data_scientist_helps_inform_and_imagine_armys_workforce_of_the_future

<u>Resources about Data Science in Chemistry</u>
Stanford AI recreates chemistry's periodic table of elements
- https://news.stanford.edu/2018/06/25/ai-recreates-chemistrys-periodic-table-elements/

GANs in Chemistry

- https://medium.com/@pagsepp/gans-in-chemistry-fe1dfce1ce85

The Advent of Generative Chemistry

- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7429972/

Recent advances and application of generative adversarial networks in drug discovery, development, and targeting

- https://www.sciencedirect.com/science/article/pii/S2667318522000150

Resources about Periodic Table

When Will We Reach the End of the Periodic Table?

- https://www.smithsonianmag.com/science-nature/when-will-we-reach-end-periodic-table-180957851/

Interactive Periodic Table of Elements

- https://pubchem.ncbi.nlm.nih.gov/periodic-table/#view=list

Ordering the Elements

- https://www.science.org/doi/10.1126/science.aav7350

Are There Undiscovered Elements Beyond the Periodic Table?

- https://www.youtube.com/watch?v=prvXCuEA1lw

How Do We Make New Elements?

- https://www.youtube.com/watch?v=Y9D1XQdIaLw

On the position of helium and neon in the Periodic Table of Elements

- https://link.springer.com/article/10.1007/s10698-017-9302-7

Resources about GANs

What are GANs?: Introducing Generative Adversarial Networks to Middle School Students

- https://ojs.aaai.org/index.php/AAAI/article/view/17821

What are GANs?

- https://www.youtube.com/watch?v=TpMIssRdhco&pp=ygUEZ2Fucw%3D%3D

How synthetic media, or deepfakes, could soon change our world

- https://www.youtube.com/watch?v=Yb1GCjmw8_8

Do Deepfakes Have Fingerprints? | Deepfake Detection + GAN Fingerprints

- https://www.youtube.com/watch?v=4lnEst2nojA

Defense Department has produced the first tools for catching deepfakes

- https://www.technologyreview.com/2018/08/07/66640/the-defense-department-has-produced-the-first-tools-for-catching-deepfakes/

Voices of DARPA Podcast, Episode 69: Demystifying Deepfakes

- https://www.youtube.com/watch?v=jUrGCCbcF60

PRE-LAB RECOMMENDATIONS

To reduce time needed to explain the mechanisms within CODAP, we recommend having students go through the following 60-minute exercise prior to the lab experience.

A Tool for Doing Data Science: CODAP

PREPARATIONS

Tablet Prep

- Load CODAP on tablets
- Open the Alien Elements file stored on the device

Computer Prep

- Load the Python Script on the computers
- Keep hidden until needed for Part 2

STUDENT STATION SET-UP

Students will work in pairs at an individual station with a tablet and will share the computer and element cards with the groups at their table.

1. Tablets with USB adapter and mouse (per station)
2. Keyboard and mouse for computer (per table)
3. Dry erase mat (per table)
4. 36 alien element cards (per table)
5. Wet erase marker (per station)
6. Pencils

LESSON PLAN

Introduction (5-10 min)

- Welcome students to the lab and direct them to take a seat.
- Explain to students that they will be acting as data scientists on a chemistry team investigating a new planet called, Dod.
- Ask students if they have heard of data science before? If not, what does it make them think of?
    - Students might know that data scientists look for information in large datasets.
    - Students might associate data with data analysis and data mining.
- Explain that data science is a way to get information and insights from large, noisy datasets.
- Ask students what data they think is collected when they are browsing online and/or interacting with social media?
    - Students might respond with search history, previous purchases, clicks on webpages, time spent looking at videos/ TikToks, and more
- Ask students why companies would want to collect this data?
    - So the companies can cater their algorithm to your location, preferences, shopping habits, etc.
- Explain that, from a business perspective, the data collected needs to be analyzed to determine what actions they can take. Perhaps a company wants to how they can best serve their audience so they look into information the data to form algorithms that help their users to see recommended products and create a personalized experience for the user.
- Explain that data science is used in many fields, not just online businesses. It can be used in any field as long as data can be collected to answer their question.
- Explain to students that data science can be descriptive, diagnostic, predictive, and prescriptive.
- Ask students what types of information we might get from a descriptive or diagnostic analysis? Ask students what does descriptive and diagnostic sound like/make them think of?
    - Descriptive allows us to understand what relationship are there and diagnostic allows us to understand why those relationships exist.
    - Students might associate descriptive with describing or detailing information and diagnostic with determining a cause (similar to medical diagnoses).
- Ask students how predictive or prescriptive analysis would be different? Ask students what does predictive or prescriptive sound like/make them think of?
    - Predictive allows us to foretell information about data that hasn't directly been collected and prescriptive allows us to use the predicted answers to create new processes/algorithms that are specific to that answer.
    - Students might predictive with seeing the future, identifying what hasn't happened yet. They might associate prescriptive with medicine and as a treatment or plan of action based on a diagnosis.
- Explain to students that we'll be starting with a descriptive analysis and will move into predictive analysis with our data today.

Part 1: Descriptive Analytics and Visualization (30 min)

- Have students turn to page 2 and reintroduce today's goal: Scientists have discovered 36 unique elements in the planet Dod's atmosphere and on its surface. The collected data has been provided to the data scientist (you!) to help communicate the relationships between the elements and their properties.
- Do a quick overview of what they can see in CODAP. *Share screen to walk them through the components.*
    - There is a table with 36 different elements and many properties including electronegativity, atomic weight, ionization energy, and more.
    - There is a text box with descriptions of each property on the right.
    - There is a graph that is filled with dots but no x or y axis label.
- Viewing data on the graph.
    - Explain how to drag properties to an axis on the graph or how to change the axis by clicking on it and selecting a new property.
    - Show that you can select points on the graph and they are highlighted in the table.
- In reference to student handout, direct students to look for a unique measured property. This will act as our dependent variable. We want something that, when graphed on the x axis, remains very flat and there are as few overlapping or stacked points as possible. *In slides, show an example of what we mean by flat. Leave this up for students to reference.*
    - Give students 5 minutes to look through the different properties.
    - Students should eventually identify atomic weight as the most unique property.
    - Remind students that we are looking for a unique **property**. If students think Alien Symbol or Alien Name are unique, remind them that these are not properties of the element, but rather human given names. These are arbitrary and though unique, are not measured properties of the element
- Next direct students to keep atomic weight as their dependent variable and add different properties to the y axis (the independent variable). Remind them to look for trends (something that always goes up or down or stays constant) and patterns (repeating increases and/or decreases).
    - Give students 5-10 minutes to look through these properties.
    - Students should identify the following:
        - **Density** trends upward with atomic weight, but falls off after B
        - **Electronegativity** creates bands showing increases with increasing atomic weight, also shows an overall trend of decreasing with atomic weight
        - **Ionization Energy** creates bands showing trends of increases with increasing atomic weight, also shows an overall trend of decreasing with atomic weight
        - **Atomic Radius** creates bands showing trends of decreases with increasing atomic weight, also shows an overall trend of increasing with atomic weight
        - **Specific Heat** shows general trend of decreasing with atomic weight
    - *Review the trends and patterns slides to show students what those graphs look like.*

- Next direct students to start their visualization. Explain to students that we want to arrange our element cards in table that helps to communicate the patterns and trends identified. Remind students that they will be working as a table of six to accomplish this.
    - Advanced groups can go through this guided procedure on their own, but to keep the class on pace it is recommended to help students through each step.
    - Give students up to 15 minutes for this part.
- Explain to students that we want to create rows that match the bands we saw for atomic radius, electronegativity and ionization energy.
- Direct students to use the graph of Atomic Radius vs Atomic Weight to sort the cards into 4 rows on your center mat. Ensure each row is ordered left to right by atomic weight.
    - *Show the slide to show what we mean by creating rows from the bands of data points.*
- After students, sort their cards into the 4 rows found in atomic radius v atomic weight, ask students if this matches the bands we identified for electronegativity and ionization energy?
    - Yes the groups are the same but the trends differ. Electronegativity has the same bands but some elements are missing as they didn't have values (A, E, C) and Ionization Energy has the same bands but with trends upward.
- Direct students to use their dry erase marker to mark the overall trends for these properties as they go across each row (atomic weight, atomic radius, electronegativity, and ionization energy).
    - They should be marked similarly to Figure 1.
    - *Show the slide to help students know where/what to label.*
        - *In the slide, compare I and A as well as Ie and It for each graph (atomic weight, electronegativity/ionization energy, atomic radius and specific heat) to see if the trends/patterns stay true for those rows.*
        -

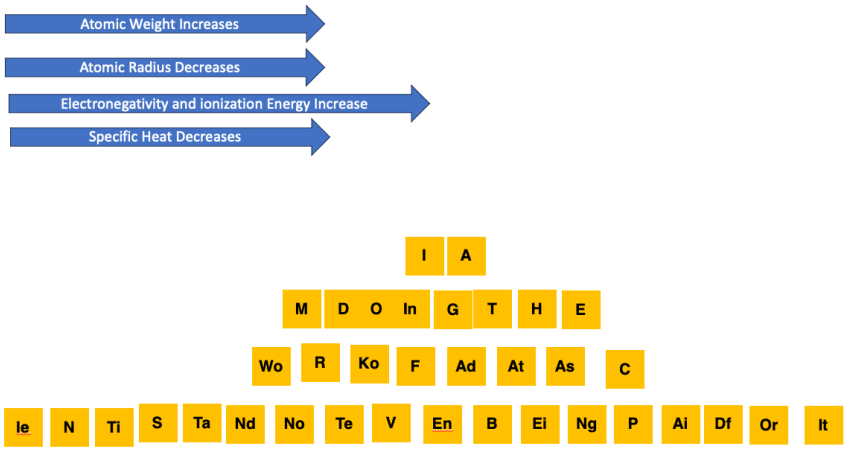

Figure 1

- Explain to students that we will now try to group the elements in columns by their chemical properties. We will use the RxOy property. Explain that this explains how many alien element atoms (R) combine with a number of oxygens to create a compound. The value is listed as a ratio of the number of R atoms (the R subscript) to the number of oxygen atoms (the O subscript).

- Remind students that as they rearrange the cards, they should make sure that the rows stay in order by atomic weights from left to right (left lower atomic weight, right higher atomic weight).
- *Use slide to show an example of what to do with the first 2:7 group.*

- At this point, the table should look similar to the one in Figure 2.

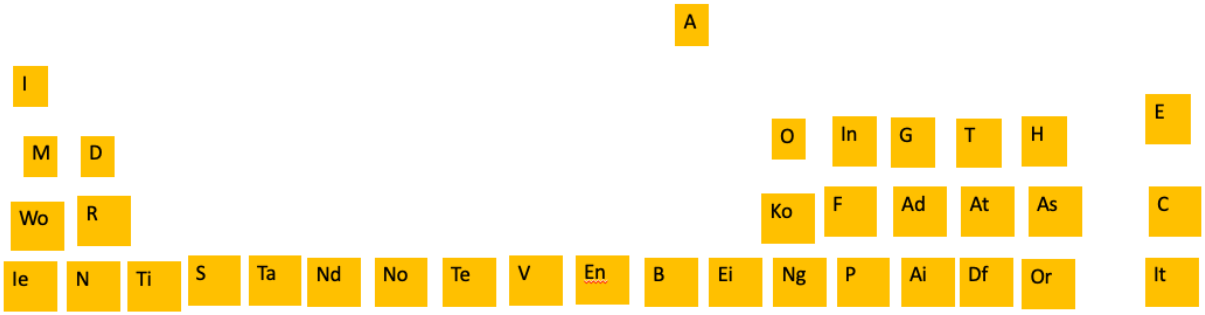

Figure 2

- If students need help determining where A, E, C and It should go, you can take a couple different paths.
  - They all are the highest atomic weight of their row so E, C, and It need to stay on the very right (18th column).
  - To figure out where A should go, you can have them look at any of the following properties and have them compare column 2 elements and column 18 elements.
    - Melting Point, Boiling Point, Density, Ionization Energy, Radius, Natural State, and Thermal Conductivity all have a trend match with group 18 (E, C, It)

- Direct students to use their dry erase marker to mark the overall trends for these properties as they go down each column (atomic weight, atomic radius, electronegativity, and ionization energy).
  - They should be marked similarly to Figure 3.
  - *Show the slide to help students know where/what to label.*
    - *In the slide, compare I and Ie as well as A and It for each graph (atomic weight, electronegativity/ionization energy, atomic radius and specific heat) to see if the trends/patterns stay true for those columns.*
    -

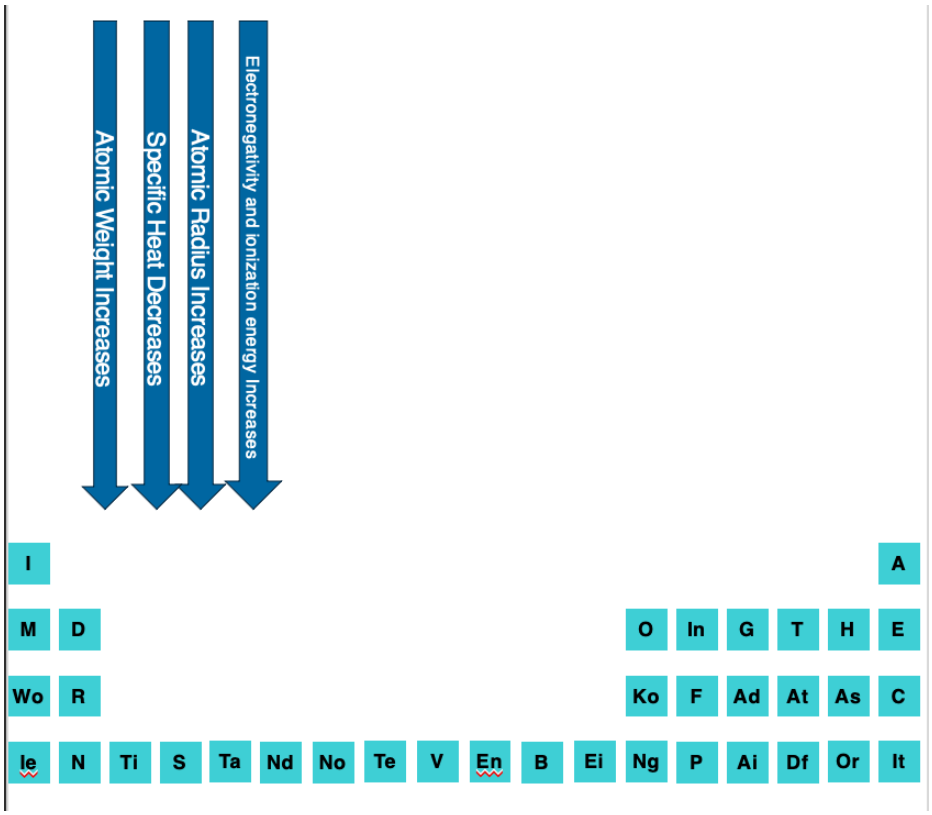

Figure 3

- Explain that we use the word periodic in the periodic table because it means that something occurs at defined intervals. Ask students if they saw any patterns of periodicity in their table?
    - Yes we see repeating patterns of reactivity with oxygen and only occurring at certain points within a row.

Part 2: Going from Descriptive to Predictive Utilizing GANs (15-20 min)

- Explain that the table we created is a descriptive visualization of the trends and patterns we saw. What we'd like to do next is use this data for a predictive analysis.
- Have students turn to page 4 and reintroduce the next goal: Scientists have found some compounds, molecules with more than one element, in the atmosphere and surface of Dod. To guide their search for more compounds, they will utilize a Generative Adversarial Network (GAN).
- Explain that a GAN uses two neural networks, a generator and a discriminator. The discriminator has been trained to recognize data, in this case the compounds found on Dod. You will work as the generator to produce new compounds that could possibly exist. But keep in mind, the discriminator is looking for the same patterns as its dataset, so you'll need to learn from your responses to find possible answers.
- Ask students, does the generator start with any information about the dataset it's trying to match?
    - No, it starts randomly and uses the discriminator's responses to find patterns.
        - This is what we call unsupervised learning because it does not require a human to help train the algorithm. The algorithm will look at the responses for patterns and will adjust its output accordingly.
- Make sure students see the note about new data. Explain that some added properties for each element have been discovered. Walk them through how to add the valence electrons to their table view.
    - *Use slide to show looped video of how to show valence electrons in the table.*
- Explain that the discriminator is hosted on the computer at each station. To start, have every group try the same compound (I_2T).
    - This compound should give the following discriminator response:
        - The compound formed is: I_2T
          Total valence electrons: 8
          Compound passed! The discriminator thinks this is a real compound!
- Direct students to record this information in their table (formula, valence electrons, pass/fail).
- Ask students where the 8 valence electrons come from?
    - Explain that because T has 6 electrons but I has 1 valence electron, but 2 atoms are present and, it all adds up to 8. We get 1 times 2 plus 6 times 1 to get 8. The subscript tells us what to multiply the element's valence electrons by.
- Ask students to make a hypothesis as to why this compound worked.
    - Nudging questions could be:
        - Is that the only ratio of subscripts that work? Does it always need to be 1 to 2? Could I do D_2Ko?
        - Is the valence electron important? Does it always need to be 8 valence electrons?
    - Allow students to try different possibilities for 10 minutes.
- Pause groups to see if any got a molecule that was accepted by the discriminator.
    - Ask for the valence electrons that was received with that compound. Ask if there is a pattern they discovered. Ask if, after hearing the class data for compounds that worked, if there is a pattern they notice.
        - The accepted molecules should have a number that is a multiple of 8.

- Once the pattern is deciphered, the generator should be able to produce any number of valid compounds. This would allow scientists to have a jumping off point when they start looking for compounds on planet Dod.

Wrap Up (5-10 minutes)

- Congratulate students on their successful first day as data scientists. Remind students that data science is a huge field with many applications that relate to their interests.
- If students are interested/if there is time, there can be a discussion about GANs.
    - We used a GAN to create synthetic data, more molecules that we do not have any physical evidence of. This can be great tool for other purposes like masking personal information and for creating larger data sets, but GANs can also be used for image, video, and sound generation. These GANs are really common in the news now because of their use to create deepfakes.
    - Ask students if they've heard of them before.
        - It allows someone to make a video that puts the image of someone else's face over their own and it looks more convincing than an ordinary filter like you might use on Instagram or Snapchat.
            - Tom Holland and Robert Downey Jr. as Marty McFly and Doc in Back to the Future.
    - Explain that these started out as a GAN and utilized its unsupervised learning feature, but the generators have become so good that they just act as generators now without a discriminator.
    - Ask students what are the dangers of deepfakes.
        - It fuels fake news and can be hard to spot. It blurs the line between reality and fiction.
    - The Department of Defense cares about this issue deeply because of the security threat it possesses. For example, there is a deepfake of Ukranian President Volodymyr Zelensky urging Ukrainian forces to lay down arms and surrender to Russia.
    - Teams at the National Security Agency (NSA) and Defense Advanced Research Projects Agency (DARPA) are working to make tools to identify synthetic media. But as these agencies develop these tools, more GANs are created to get around them.