# Transformer-based DNN for Identification of RFI Sources in Astronomical Observations

Arush S. Sharma and Marwan Krunz
Department of Electrical and Computer Engineering, University of Arizona, Tucson, USA
{arushsharma, krunz}@arizona.edu

***Abstract*—Radio astronomical observations provide valuable insights into celestial phenomena such as solar systems, cosmic galaxies, and other astrophysical sources. Using highly sensitive telescopes, these observations typically capture faint cosmic signals at extremely low power levels. However, such sensitivity also makes astronomical data susceptible to contamination by unwanted man-made signals, collectively known as Radio Frequency Interference (RFI). RFI commonly arises from satellite transmissions, cellular networks, automotive radar systems, and others. Identifying the sources of RFI is therefore critical for the radio astronomy community, as it can guide the development of interference mitigation techniques and passive/active shared spectrum policies. In this paper, we utilize RFI observations from the single-dish Green Bank Telescope (GBT) in Green Bank, West Virginia. Specifically, we utilize measurements collected between 2020 and 2022 over the L, S, and C frequency bands. Sub-bands within these bands are labeled using the GBT-RFI-GUI tool, which annotates frequency segments according to Federal Communications Commission (FCC) allocation tables. For each of the three bands, we train a Transformer-based neural network to classify the RFI sources into the appropriate sub-bands. Our classifier achieves inference accuracy of 84%, 87%, and 93% for the L, S, and C bands, respectively. Furthermore, to enhance the classifier's performance, we design and train Generative Adversarial Networks (GANs) and use these GANs to synthesize additional data samples for sub-bands with scarce training data. We observe that when the classifier is trained on an augmented dataset that combines both synthetic and real samples, its inferency accuracy is improved to 91%, 98%, and 97% for L, S, and C bands, respectively.**

***Index Terms*—RFI, Generative AI, GANs, radioastronomy, Transformer, Deep learning**

## I. Introduction

Radio astronomy (RA) telescopes play a vital role in studying Earth's environment, our solar system, distant galaxies, and celestial sources such as pulsars and stars. Operating across frequencies from 2 MHz to beyond 1000 GHz, these telescopes capture a wide variety of astrophysical emissions.

With the ever-increasing demand for wireless capacity, telecommunications systems are expanding their operation into higher frequency ranges, including those above 100 GHz, with the goal of leveraging tens of GHz of unallocated or partially allocated spectrum. The so-called sub-terahertz (sub-THz) spectrum, once deemed unsuitable for wireless communications due to severe propagation losses, is becoming viable owing to advances in directional antennas and beamforming technologies. Consequently, spectrum sharing between active communication systems and passive scientific services, such as radio telescopes and weather satellites, has become an important research topic.

RA telescopes are designed to detect extremely weak cosmic emissions, often orders of magnitude below man-made signals. Technological advances in adaptive filtering and signal processing have further increased this sensitivity. A major challenge for the radio astronomy community is the persistent presence of Radio Frequency Interference (RFI) from man-made emitters, which can render RA observations unusable and necessitates costly remeasurements. For shared bands, RFI occupying more than 2% of the observation period can significantly degrade data quality. To mitigate RFI, most observatories are located in remote regions; for e.g., the National Radio Quiet Zone (NRQZ) in the Green Bank Telescope (GBT) [1] restricts radio emissions within its vicinity.

Despite their relative isolation from man-made signals, telescopes still suffer from RFI, particularly in the lower frequency bands (e.g., L, S, and C), where communication satellites, terrestrial transmitters, etc., operate. At the GBT, dedicated spectrum-monitoring campaigns are routinely conducted to characterize this interference environment. These measurements are typically performed during non-observing periods, when the telescope is used exclusively to record RFI activity, ensuring that the resulting scans capture only man-made emissions rather than astronomical signals.

Previous studies primarily focused on *binary* RFI detection, distinguishing astronomical signals from interference. Such coarse classification is insufficient for effective spectrum management or interference mitigation. In contrast, identifying which type of RFI source contaminates a radio astronomical observation (e.g., satellite, radar, or cellular) can provide deeper diagnostic insights, guide the development of source-specific mitigation strategies, and inform regulatory spectrum coordination. To the best of our knowledge, this work presents one of the first efforts to classify multiple RFI sources from real telescope observations using deep learning.

In this work, we leverage Transformer-based neural networks to learn complex temporal–spectral characteristics of RFI within the L, S, and C bands observed at the GBT. To address data scarcity in certain frequency sub-bands of these three bands, we employ Generative Adversarial Networks (GANs) for generating synthetic samples. The synthetic data is used to augment the original data and retrain the classifier, leading to improved classification accuracy. The main contributions of this paper are as follows:

979-8-3315-7111-5/26/$31.00 ©2026 IEEE

- We construct a comprehensive RFI dataset using spectrum monitoring observations from the GBT collected between 2020 and 2022 and covering the L, S, and C frequency bands. The sub-bands are labeled using the GBT-RFI-GUI software, which annotates frequency segments based on FCC allocations, thereby providing a reliable indication of interference sources.
- We develop a Transformer-based neural network for multi-class RFI source classification. Unlike binary RFI detection (also known as flagging), the proposed model distinguishes between different types of interference and accommodates variable-length input sequences corresponding to different sub-band bandwidths. The trained model achieves test accuracies of 84%, 87%, and 93% for the L, S, and C bands, respectively.
- To address data imbalance and scarcity in certain frequency sub-bands, we employ GANs to synthesize additional training samples. Augmenting the training dataset with synthetic samples results in a notable improvement in the overall classification accuracy up to 91%, 98%, and 97% for L, S, and C bands, respectively.

## II. Related Work

Several previous works introduced various techniques for detecting and identifying RFI. Mesarick et al. [2] identified RFI by training machine learning (ML) models on clean data and using these models to separate astronomical signals from interference. Their method employs nearest latent neighbors, leveraging distances in the autoencoder's latent space for detection. The performance was evaluated on simulated Hydrogen Epoch of Reionization Array (HERA) data and real Low-Frequency Array (LOFAR) observations using AUROC, AUPRC, and maximum F1-score. Lourenco et al. [3] surveyed and monitored RFI within the Australian Square Kilometre Array Pathfinder (ASKAP) 700–1800 MHz range, archiving flagging statistics for near real-time RFI tracking. By using the telescope as a sensitive RFI detector, they determined that around 3% of scientific data is affected by interference, primarily from mobile (742–1085 MHz), aeronautical (1293–1437 MHz), and satellite navigation services (1510–1797 MHz). Li et al. [4] introduced Allspark-UNet, a semi-supervised segmentation model that reduces dependence on labeled datasets. The model incorporates a feature enhancement module to better reconstruct the spectral–temporal RFI structure, achieving good performance with limited annotations. Tested on real data from the 40-meter Yunnan Observatory telescope, the model gave 98% accuracy of RFI detection using only 272 labeled samples, surpassing baseline methods by 1.52% in F1 score and 2.18% in mean intersection over union.

Unlike previous works, which primarily focus on RFI detection (presence or absence of RFI), our study addresses the classification of RFI type, aiming to identify and categorize distinct interference emitters rather than merely detect their presence.

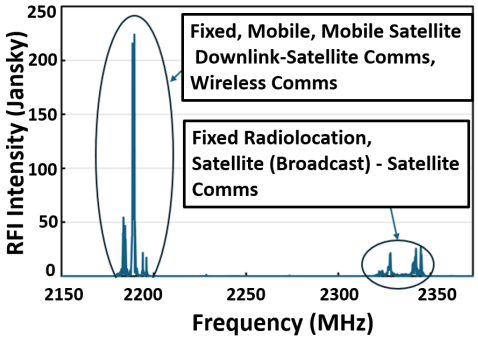


Fig. 1. RFI in the S-band.

## III. Astronomical GBT Dataset

GBT is a 100-meter fully steerable single-dish telescope with a frequency range from 290 MHz to over 116 GHz [5]. It is situated within the NRQZ, which provides protection against harmful RFI. RFI scans are routinely captured by telescope operators during non-scientific observation periods to monitor the surrounding radio environment. These scans are typically obtained by rotating the telescope 360 degrees in azimuth, capturing the full spectral profile of ambient emissions.

We utilize the 'GBT-RFI-GUI' software [6] to access RFI data stored on the GBT servers. The software allows users to retrieve scans by specifying the desired receiver and the observation time interval and visualizing the spectral data as a plot of RFI intensity (in Janskys) versus frequency. Figure 1 shows one such example. Additionally, the GUI displays annotations for individual sub-bands within each scan, where the frequency allocation corresponds to FCC designations [7].

Our analysis focuses on the L, S, and C bands, which are heavily used by various wireless technologies. The dataset covers observations from 2020–2022. Tables I, II, and III list the corresponding sub-band ranges and bandwidths along with the number of samples for each sub-band. Each sample represents a vector of RFI intensity at various frequencies. The frequency resolution is 1.4 kHz in the L band, 91.43 kHz in the S band, and 65.87 kHz in the C band. Many sub-bands contain multiple licensed transmitters, each underscoring the dense interference environment. Sub-band identifiers follow the labels defined in Section IV.

## IV. RFI Classification

We design a Transformer-based [8] neural network tailored for classifying different sources of RFI in the GBT observations. The model is constructed to handle variable-length spectral segments and capture long-range dependencies within frequency sequences, which is difficult to deal with using conventional convolutional or recurrent architectures.

The input to the model is a 2D sequence of dimension $(L \times D)$, where $L$ is the chunk size and $D$ is the feature dimension. A chunk corresponds to a contiguous segment of the original sequence, and in our configuration, its length is fixed at 64. Each input sequence first passes through a linear projection layer that maps it to an embedding space. To

TABLE I
RFI IN THE L-BAND AT THE GBT

| Sub-band Label | Sub-band Name | No. of Samples | Frequency Range (MHz) | Bandwidth (MHz) |
|---|---|---|---|---|
| L1 | Aeronautical Mobile/Radionav - Aviation | 167 | 1107-1160 | 53 |
| L2 | Earth Exploration, Radiolocation, Sat-Nav, Space Research | 17 | 1220-1240 | 20 |
| L3 | Amateur & Scientific SatCom | 125 | 1240-1300 | 60 |
| L4 | Aeronautical Radionav, Radiolocation - Aviation | 98 | 1300-1350 | 50 |
| L5 | Fixed, Mobile Radiolocation | 12 | 1350-1390 | 40 |
| L6 | Mobile (Aeronautical Telemetry) - Aviation | 140 | 1435-1525 | 90 |
| L7 | Mobile - Satellite Downlink, Maritime | 161 | 1525-1535 | 10 |
| L8 | Mobile - Satellite Downlink, Maritime, Aviation | 168 | 1535-1559 | 24 |
| L9 | Satellite Aviation services | 173 | 1613.8-1626.5 | 12.7 |
| L10 | Mobile - Satellite Uplink, Maritime, Aviation | 20 | 1626.5-1660 | 33.5 |
| L11 | Meteorological Observations | 30 | 1675-1695 | 20 |
| L12 | Meteorological Sat & Terrestrial Wireless | 17 | 1695-1710 | 15 |

TABLE II
RFI IN THE S-BAND AT THE GBT

| Sub-band Label | Sub-band Name | No. of Samples | Frequency Range (MHz) | Bandwidth (MHz) |
|---|---|---|---|---|
| S1 | Fixed, Mobile, Space Operation (Uplink) | 2 | 1780-1850 | 70 |
| S2 | Fixed/Mobile RF & Microwave Services | 13 | 1850-2000 | 150 |
| S3 | Terrestrial & Mobile-Satellite Comms | 8 | 2000-2020 | 20 |
| S4 | Space Operations & TV Broadcast Services | 9 | 2025-2110 | 85 |
| S5 | Fixed Mobile | 6 | 2120-2180 | 60 |
| S6 | Fixed, Mobile, Mobile Satellite Downlink, Wireless Comms | 119 | 2180-2200 | 20 |
| S7 | Space Operations & Line-of-Sight Telemetry | 10 | 2200-2290 | 90 |
| S8 | Fixed/Mobile Wireless & Satellite Broadcasting | 69 | 2310-2320 | 10 |
| S9 | Fixed Radiolocation, Broadcasting Satellite Comms | 117 | 2320-2345 | 25 |
| S10 | Fixed/Mobile Wireless & Satellite-Broadcasting | 74 | 2345-2360 | 15 |
| S11 | Fixed/Mobile RF (Non-Aero) | 5 | 2500-2655 | 155 |

TABLE III
RFI IN THE C-BAND AT THE GBT

| Sub-band Label | Sub-band Name | No. of Samples | Frequency Range (MHz) | Bandwidth (MHz) |
|---|---|---|---|---|
| C1 | Land-Based Wireless Communications | 25 | 3869-4000 | 131 |
| C2 | Fixed, Fixed Satellite Downlink - Satellite Comms | 37 | 4000-4200 | 200 |
| C3 | Aeronautical Radionavigation - Aviation | 14 | 4200-4400 | 200 |
| C4 | Fixed, Mobile | 7 | 4400-4500 | 100 |
| C5 | Fixed/Mobile + Fixed-Sat & TV Relay Services | 8 | 6875-7025 | 150 |
| C6 | Fixed RF & Space Research (Deep Space) | 18 | 7145-7190 | 45 |
| C7 | Earth Exploration & Space Research (Uplink RF) | 6 | 7190-7235 | 45 |
| C8 | Earth Exploration Satellite Uplink, Fixed - RF Devices | 12 | 7235-7250 | 15 |
| C9 | Fixed & Mobile Satellite Uplink, Fixed | 17 | 7250-7300 | 50 |
| C10 | Fixed, Fixed & Mobile Satellite Downlink | 22 | 7300-7375 | 75 |
| C11 | Multi-Service Satellite Downlinks & Terrestrial Wireless | 25 | 7375-7450 | 75 |
| C12 | Multi-Satellite Downlink & Fixed Services | 23 | 7450-7503 | 53 |
| C13 | Fixed Services & Multi-Satellite Downlinks | 64 | 7500-7750 | 250 |
| C14 | Fixed, Meteorological Satellite Downlink | 69 | 7750-7900 | 150 |

maintain the spectral ordering across sub-bands, we introduce a global positional encoding scheme that assigns positions according to their absolute frequency location rather than resetting them within each chunk.

The encoded sequence is then processed through a stack of Transformer encoder layers, each comprising a multi-head self-attention block followed by a position-wise feed-forward network. Residual connections and layer normalization are applied after both sub-blocks to stabilize training and enhance gradient flow. The output of the final encoder layer is aggregated using global average pooling across the sequence dimension, producing a compact representation of the entire sub-band. This representation is passed through a dense classification layer with a softmax activation to predict one of $N$ classes of interference.

For our implementation, we use four attention heads and two encoder layers. The data is split into 80/20 for training and testing, respectively. The model is trained for 200 epochs using the Adam optimizer and a sparse categorical cross-entropy loss function with a batch size of 64.

Figure 2 depicts the architecture of the classifier. Figures 3, 4, and 5 show the confusion matrices for L-band, S-band,

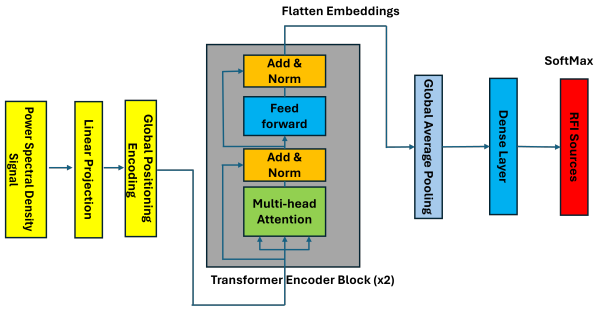


Fig. 2. Transformer-based RFI classifier.

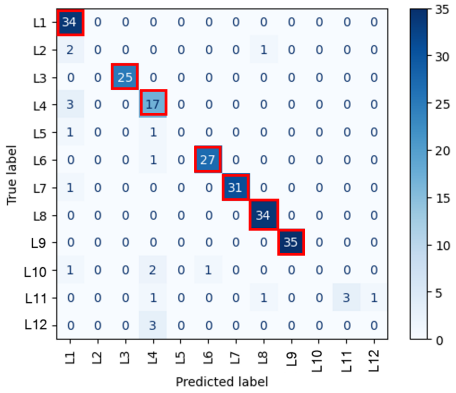


Fig. 3. Confusion matrix for the L-band classifier.

and C-band classifiers, respectively. The cells highlighted in red denote sub-bands with sufficient test samples to provide statistically robust estimates of the available training data. For the L-band, the classifier accurately identifies the RFI source in sub-bands L1, L3, L4, L6, L7, L8, and L9. Its performance for the remaining L sub-bands is less impressive, likely due to the small number of available training samples. Similar trends are observed for the S-band, where the classifier achieves high accuracy for S6, S8, S9, and S10. The remaining sub-bands have very few samples to provide reliable statistics. The classifier on the C-band data shows higher accuracy for C13

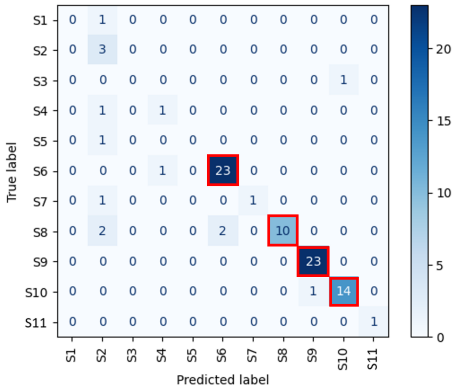


Fig. 4. Confusion matrix for the S-band classifier.

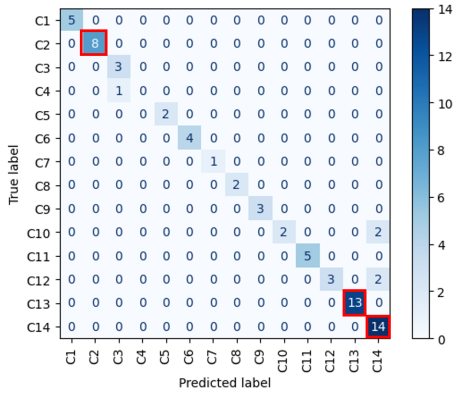


Fig. 5. Confusion matrix for the C-band classifier.

and C14 sub-bands, although the other sub-bands have too few examples to be considered statistically stable. Overall, the test accuracy for L-band, S-band, and C-band classifiers is 84%, 87%, and 93%, respectively.

## V. Data Augmentation Using Generative AI

Generative AI techniques are valuable for ML classifiers when data is scarce, as intermittent RFI sources around telescopes are infrequently captured. This hampers the evaluation of the classifier's performance. In this section, we focus on augmenting the data of specific sub-bands of L, S, and C-bands using GANs.

The GAN model was proposed by Goodfellow et al. [9] in 2014 for unsupervised learning. In its basic form, the model consists of two components: generator and discriminator. Using a game theoretic approach [10], the generator tries to learn the underlying distribution of the real data and generate synthetic samples that are statistically similar to the real data. The discriminator, on the other hand, tries to distinguish the real samples from the synthetically generated ones. Both the generator and discriminator are neural networks that are jointly trained. As shown in Figure 6, we use a convolutional neural network (CNN) comprising 5 convolutional layers (CL) for the discriminator.

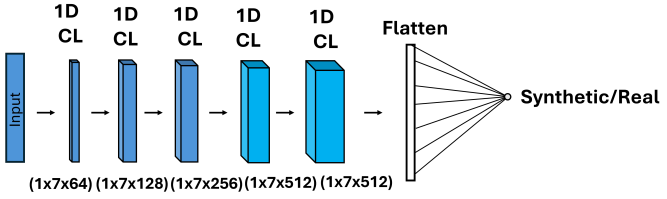


Fig. 6. Architecture of a discriminator in the GAN model.

As for the generator, its input consists of a 128-long noise vector and is composed of successive convolutional layers. However, the number of layers depends on the output size, which varies from one sub-band to another. Figure 7 depicts an example generator for the sub-band S3. Both the generator and discriminator are trained for 200 epochs.

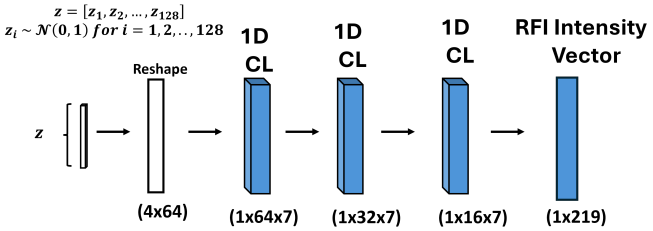


Fig. 7. Architecture of a generator for S3 sub-band in the GAN model.

### A. RFI Analysis

Using GANs, we synthesize 5 L sub-bands (L2, L5, L10, L11, L12), 7 S sub-bands (S1, S2, S3, S4, S5, S7, and S11), and all 12 C sub-bands. Synthesis is done to make sure that the dataset is balanced and the results are statistically valid. Subsequently, we split the data into 80/20 for training and testing and then train the Transformer-based classifier on the augmented dataset. Figures 8–10 show the confusion matrices

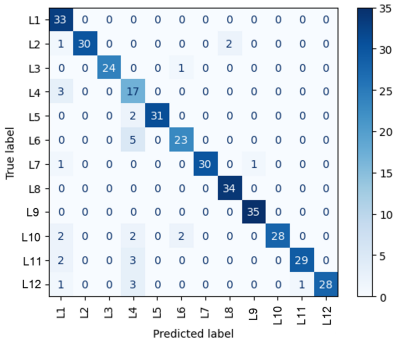


Fig. 8. Confusion matrix for the L-band classifier after synthesis.

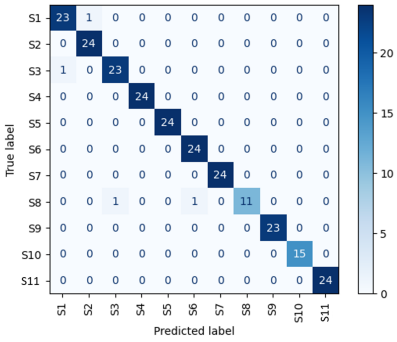


Fig. 9. Confusion matrix for the S-band classifier after synthesis.

for representative L, S, and C sub-bands. Synthesized data improves performance across all cases, yielding accuracies of 91%, 98%, and 97% for the L, S, and C bands, respectively.

## VI. Conclusions

In this paper, we used the RFI scans from the GBT telescope covering L, S, and C bands from 2020 to 2022 to design a Transformer-based classifier of RFI type. The sub-bands in

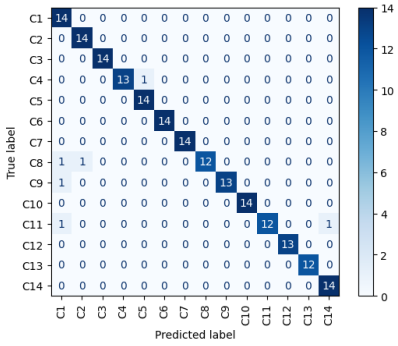


Fig. 10. Confusion matrix for the C-band classifier after synthesis.

these three bands were annotated by leveraging the 'GBT-RFI-GUI' software, which assigns labels to these sub-bands based on the FCC allocation tables. For each band, the classifier was trained separately, and the results were reported in terms of confusion matrices where performance across the L, S, and C bands was summarized. Finally, to overcome data scarcity in specific sub-bands, the generative AI technique was leveraged, where the classifier's performance was enhanced by generating synthetic samples using GANs that resulted in inference accuracy of 91%, 98%, and 97% for L, S, and C bands, respectively. Future work may explore expanding the dataset to additional frequency bands and incorporating more diverse RFI environments to further improve model robustness.

## References

[1] NRAO, "National Radio Quiet Zone (NRQZ)." https://info.nrao.edu/do/spectrum-management/national-radio-quiet-zone-nrqz-1. Accessed: Nov. 22, 2025.

[2] M. Mesarcik, A.-J. Boonstra, E. Ranguelova, and R. V. van Nieuwpoort, "Learning to detect radio frequency interference in radio astronomy without seeing it," *Monthly Notices of the Royal Astronomical Society*, vol. 516, no. 4, pp. 5367–5378, 2022.

[3] L. Lourenço, A. Chippendale, B. Indermuehle, V. Moss, T. Murphy, T. Galvin, G. Hellbourg, A. Hotan, E. Lenc, and M. Whiting, "Survey and monitoring of ASKAP's RFI environment and trends I: Flagging statistics," *Publications of the Astronomical Society of Australia*, vol. 41, p. e012, 2024.

[4] J. Li, B. Liang, S. Feng, W. Dai, and S. Wei, "RFI detection based on semi-supervised learning with improved Unet," *Astronomy and Computing*, vol. 54, p. 101020, 2026.

[5] A. Schmiedeke, "The GBT." https://gbtdocs.readthedocs.io/en/latest/references/gbt.html. Accessed: Nov. 22, 2025.

[6] Green Bank Observatory, "RFI GUI User Guide." https://greenbankobservatory.org/about/interference-protection/rfi-gui-user-guide/. Accessed: Nov. 22, 2025.

[7] Federal Communications Commission, "FCC online table of frequency allocations." https://transition.fcc.gov/oet/spectrum/table/fcctable.pdf, 2022. Accessed: Nov. 22, 2025.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. of the Advances in NeurIPS Conference*, vol. 30, 2017.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. of the Advances in NeurIPS Conference*, vol. 27, 2014.

[10] M. Mohebbi Moghaddam, B. Boroomand, M. Jalali, A. Zareian, A. Daeijavad, M. H. Manshaei, and M. Krunz, "Games of GANs: game-theoretical models for generative adversarial networks," *Artificial Intelligence Review*, vol. 56, no. 9, pp. 9771–9807, 2023.